THE USE OF DIGITAL COMPUTERS FOR STATISTICAL ANALYSIS IN TEXTILES

A THESIS

Presented to

The Faculty of The Division

of Graduate Studies

By

Sharad R. Sarvate

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Textiles

Georgia Institute of Technology

March 1978

THE USE OF DIGITAL COMPUTERS

FOR STATISTICAL ANALYSIS IN TEXTILES

Approved:

Milos Konopasek

David R. Gentry

L. Howard Olson

Date approved by Chairman 12/10/7

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to my thesis advisor and friend, Dr. Milos Konopasek, whose guidance and encouragement made this thesis possible.

Special thanks are due to Dr. David R. Gentry and Dr. Howard Olson for their valuable advice and for serving on my reading committee.

I am very thankful to the Graduate Division for allowing certain deviations from normal format of the thesis in order to print it with the help of the Georgia Tech CYBER '74 Computer.

Miss Mary Trauner and Mr. John Pearson of the Office of Computing Service (OCS) rendered invaluable assistance with the computer work for which I am very grateful.

I am also thankful to Dr. W.D.Freeston, for providing partial financial assistance.

Finally, I want to thank other faculty members and friends for their suggestions and help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

There is a long tradition of using statistical methods for analysis of data in textile quality control and research. Little use is made, however, of digital computers in implementing the more sophisticated, powerful and time saving statistical methods.

In this thesis the current status of statistical methods used in textiles has been reviewed as is reflected in the survey of current literature and research activities at the School of Textile Engineering.

Many of the major areas of statistical analysis are covered in this report, including measures of central value and dispersion, regression and correlation, and hypotheses testing. Special attention has been paid to the man-computer interaction, and to the relative merits of using available program packages, in particular the Statistical Package for Social Sciences (SPSS). Selected methods have been illustrated by giving examples of their application to practical problems in textiles. Steps for using computers for statistical analysis have been exemplified. The implementation of new and fast developing calculation devices like programmable calculators and microcomputers in statistical analysis is also discussed.

CHAPTER I

INTRODUCTION

1.1 Statement Of The Problem

Textile quality control and research involves extensive experimental work and collection of large amount of data. Analysis of data and decision making is usually very time consuming. Also the results cannot be fully exploited if the analysis is not exhaustive and accurate.

A good knowledge of statistical methods is required to draw the correct conclusions from test results. This work could be tedious for those who are not well acquainted with statistics. Even with knowledge of statistical methods, it is difficult to do the calculations using only desk or pocket calculators. Computers can simplify the mathematical work but developing computer programs for complicated statistical procedures is not within the capability of many quality control managers and textile researchers.

This thesis describes the most often used statistical methods in a simplified manner and suggests the easiest way to use a computer for handling them. In addition to the usual descriptive statistics, simple frequency distributions and histograms, this report contains procedures for simple correlation, partial correlation, means and variances for subpopulations, oneway and n-way analysis of variance, multiple regression, scatter diagrams, and factorial analysis.

The intellectual process of inductive research ideally goes thru the following stages [16] :

1. The researcher begins with a set of ideas concerning the operation of certain aspects of technical reality. This involves the isolation of variables at the conceptual level and the formation of some general notions concerning their interrelationships and causal effect upon each other.

2. An empirical data base is generated, containing indicators of the conceptual variables in which the researcher is interested.

3. The researcher then formulates more concrete hypotheses concerning what pattern of interrelationships should be found in the empirical indicators if the original ideas about the operation of technical reality were correct.

4. The data are then analyzed using one or more of a variety of statistical techniques in order to determine whether the expected pattern of relationships can actually be discerned in the data.

Most often it is then discovered that at best the actual patterns in the data only partially reflect the original concepts. Conceptions are altered in light of the empirical findings and further analysis is performed to test these ideas. The data suggest new ideas, which in turn suggest new analysis. This iterative process continues in the hope that the researcher will be able to reach an understanding of interrelationships and cause and effect as they are reflected in the pattern of the data. It is then obvious that faster computation methods provide greater scope to draw conclusions.

The flow of large amounts of information and extensive information processing is characterstic for statistical analyses and the use of computers as powerful man-made information processing devices is an obvious choice. The main objective of this thesis is to show how the computer can be used efficiently for statistical calculations in textiles at present time and in an environment serviced by a large computer like Georgia Tech's Cyber 74.

Thus, the problem we attempted to solve during the work on this thesis was two fold: (1) to summarize the state of the art in using statistical methods in textile research and qualty control and, (2) to introduce textile scientist or engineer to the use of the computer for this purpose. The rest of this chapter deals with the first item and most of the remaining part of the thesis deals with the use of the

"Statistical Package for Social Sciences" (SPSS) in statistical computation, which, as will be explained later, seems to be the best tool for given purpose [16].

1.2 Review of literature

Statistics is the art and science of gathering, analyzing and making inferences from data. The foundations of mathematical statistics have been laid in 19th and first half of 20th century by Gauss (method of least squares), Fisher (design of experiments), Wald (quality control) and others. Progress in science, technology and government control stimulated the development of a large variety of statistical theories and methods which are widely used in experimentation, decision making and control in practically every field of human endeavor. The statistical methods can be classified as follows [44,45] :

1. The analysis of statistical data with the help of frequeny distributions, histograms, measures of location, measures of variation, skewness and kurtosis.

2. The probability and probability distribution.

3. Sampling and sampling distribution like normal distribution, Student's T - distribution, chi-square distribution and Fisher's F - distribution.

4. Statistical estimation for one-sided and two-sided confidence intervals.

5. Statistical hypothesis testing by establishing a decision rule, testing hypotheses about means, testing hypotheses about the proportions and testing variances.

6. Regression analysis and correlation which involves scatter diagram, linear curve fitting and curvilinear curve fitting.

7. Analysis of variance for testing difference of several means, two variables classification and additivity in the two-variable-classification model.

8. Nonparametric statistics related to contingency table, the sign test, rank test, the Kolmogorov-Smirnov statistics and Corner test for association.

A review of statistical methods used in textile quality control and research reveals that different statistical methods are used to varing degrees as shown below :

1. Commonly used methods : Mean, standard deviation, coefficient of variation.

2. Occasionally used methods : Student's T-test for comparison of two sets of data, F-test for detecting difference between two standard deviations, analysis of variance and regression analysis.

3. Rarely used methods : Factorial analysis of variance and chi-square tests. The application of these two important methods has been described in this report with illustrations.

There are several textbooks and monographs concerned entirely or partially with applications of mathematical statistics to textiles, as listed below :

1. *Physical Methods of Textile Testing* [43] by J.E. Booth. It explains only the basic statistics viz mean, standard deviation, coefficient of variation, confidence intervals, T-test and F-test.

2. An Outline of Statistical Methods used in the Textile Industry [3] by A. Brearley. This covers a wide range of statistical methods used in textile

quality control. In addtion to the above mentioned statistical methods it also describes analysis of variance and simple linear, multiple linear and curvilinear regression analysis.

3. Introduction to Physical Textile Testing by W. Waters [22]. Elementary statistical analysis such as measures of central tendencies, measures of dispersion and usefulness of control charts have been discussed in this book.

4. Statistical Methods for Textile Technologists by T. Murphy, et al. [24]. It includes introduction to statistical reasoning in textile work and is intended to be a practical handbook of statistical methods for use in the day-to-day work of testing and experimenting in the mill.

Many general and special methods were described in periodical literature, usually in connection with particular technical and research problems to which they were applied.

Appropriate statistical techniques useful in the research laboratory were presented in a concise form by A. Brearley in 1949 [3]. In 1959, Enrick and Lawrence described the effective use of basic statistical methods like mean, standard deviation and F-test for evaluation of significance of difference in two sets of data at various stages in textiles [5]. In 1962, Enrick also described the application of Variation Flow Analysis (VFA) for quality control [6]. Reduced product variability, improved uniformity and strength, fewer ends down, better weaveroom performance and improved appearance of yarn and fabric are some of the advantages of VFA.

The first approach to data analysis is to compute the central values and variation [6,7]. This further helps in evaluation of the significance of a difference in two test averages [5].

Little has described a unique method of detecting abnormal breakage rate in spinning [15]. It is observed that the frequency of end breaks in spinning follows a Poisson distribution if no systemetic source of variation is present. Therefore out-of-control end breaks can be detected by comparing actual distribution of breaks with expected Poisson distribution by chi-square test. The Southern Regional Research Center of the U.S. Department of Agriculture has prepared a FORTRAN program applicable to an IBM 1620 digital computer. It can be used to generate the Poisson distribution of any number of spindles on a spinning frame for total number of breaks ranging from 1 thru 1000 [15].

Borwankar has described essential requirements for a good design of experiment [2]. One of the obvious requirements is the randomization of experiment.

Randomization makes a test of significance valid by eliminating bias, and making it appropriate to analyze the data. Also the importance of replication is discussed.

When the variance of a group of items before treatment is to be compared with the variance of the same group after the treatment, it is necessary that random character of sampling is preserved. The necessity of randomized design of experiments is explained by Cochran [39] in his paper on "Testing two correlated variances".

Duckworth and Wyatt [20] have explained quick tests for the population mean and the comparision of two population means. A rapid method for analysis is also described. Ways of predicting the number of trials required in certain circumstances are given.

Mandel and Lashof have presented a new approach for the analysis of interlaboratoy studies of test methods [21]. Various sources of variability in test methods are first reexamined and a new general scheme to account for them is proposed.

Duncan [25], Enrick [24], Peterson [27] and Box [28] have illustrated quality control program and industrial statistics, containing a simplified set of procedures for use of new techniques of obtaining an optimum combination of variables for accomplishing a desired end product, such as highest quality, greatest yield or minimum waste.

Vaidya [30] has described statistical methods for the textile application of decision making in his paper entitled "Statistical quality control as a management technique. The study distinguishes between random and systematic variation and gives statistical techniques to use for valid interpretation of end breakage data in spinning and weaving processes.

Burrows [33] has explained methods of setting tolerance limits under different assumtions, and tables for this purpose are provided. This is very widely used in preparation of control charts for textile processes.

Goodman [36] and Bartholomew [41] have suggested methods for testing null hypothesis when sample is treated as a random variable. Methods for obtaining simultaneous confidence intervals for estimating the magnitude of the three factor interaction in a three-way contingency tables are developed.

A least square interactive method of fitting data to a model is described by Verbeck [18] in his work on the determination of spinning quality parameters. An actual spinning quality parameter analysis is described in datail and all results relating fiber and yarn properties are tabulated.

1.3 Justification of the Thesis

The statistical methods used in this School of Textile Engineering were surveyed in order to evaluate how the computer could be used to simplify and speed up the process of data analysis. Sixty two thesis of textile graduates from past 9 years were studied. It was found that 24 of them had used statistical methods for the analysis of experimental data. Most of the researchers had used only basic statistics like mean, variance and coefficient of variation [46 thru 54]. Vague terms such as "slightly better", "very less" were used for comparing the results. It would have been more appropriate to use t-test for comparison of data and show the statistical significance level of the difference between the results. The process of statistical significance also eliminates the difference due to chance fluctuations.

Five researchers had used the analysis of variance for decision making [55 thru 59]. Mauldin had repeated analysis of variance four times to compare the effect of dyeing processes on color difference [55]. This could have been done by SPSS in one step. Four had used regression analysis for predicting the relationship between different parameters [61 thru 65]. Bates [66] and Ko [61] not only formulated the curvilinear regressions but also evaluated the validity of the coefficients. Pulis [62], and Sakai [64] have not tested the level of significance of regression coefficients. The analysis of variance and regression analysis is very time consuming and the same can be done more quickly and accurately with the help of SPSS.

Syen [63] used linear regression for modeling an influence of twist multiplier and combing roller speed on energy consumtion, yarn strength and yarn extension in open end spinning. Neither standard error of the regression coefficients was evaluated, nor any attempt was made to test the significance level of the line of regression. In fact, according to the illustrations some of the experimental data strongly indicate nonlinear relationships and some of them do not indicate any functional relationship at all.

Coff [71] used regression analysis for establishing a relationship between panel areas, waists and in-seam lengths of trousers based on incomplete data concerning panel areas in markers with different combinations of sizes. The use of SPSS would speed up the procedure (see chapter VI).

A comprehensive survey of the use of statistical methods in the above mentioned 24 theses is given in table 1 with comments concerning possible improvements. If the researchers were using SPSS, they would save much of their time and effort and at the same time they would exploit much more information generated in the process of experimentation. In some cases their conclusions could have been firmer, in some other cases they could have been saved from questionable conclusions or they could have enhanced the credibility and validity of their results by additional experiments.

There are understandable reasons why they did not use computer more extensively and why in particular they did not use SPSS or similar package. Early in the surveyed period convenient computing facilities or SPSS were not available. In recent years the major obstacle was the amount of knowledge and degree of involvement necessary for using the computer; this was incompatible with the researchers' background and workload. It applies not only to writing original programs in FORTRAN but also to the use of SPSS which was supposed to ease the programming burden.

The fact that the most common statistical methods along with computation procedures are explained in the SPSS Manual does not encourage or help the potential user. The disadvantages of the manual are as follows:

1. It is voluminous (660 pages). It can not be used by parts as a reference material on particular subject of interest. In fact, it makes no sense unless studied from the very first page to the last.

2. It is not possible for a casual user to study the whole manual and master the techniques. It is too time consuming. A user would prefer to analyze the data manually to using SPSS if he has to spend a lot

of time learning it.

3. SPSS Manual contains too many details which are not essential for textile research and quality control

4. Brief description of statistical methods precedes its SPSS computational procedure. But these statistical methods are not illustrated and are of little value to the beginner.

In contrast to the above disadvantages of SPSS Manual, the main features of this report are :

1. This report intends to be short and precise.

2. All the unnecessary details of SPSS procedure are avoided in order to make it simple and easy to understand.

3. Most of the statistical methods required for research and quality control are included in this report.

4. Statistical methods are explained, illustrated and SPSS procedure for the same illustrations is presented in the Appendices.

This all should justify a recommendation to use SPSS as well as the approach to presenting its features to a potential user. However, we must mention for the sake of fairness that SPSS is not the last word in mathematical statistics oriented program packages. In fact, there are quite a few in existence already and many others to come. We have selected the SPSS because it is fairly comprehensive and powerful, and because it is availabe at the Georgia Tech's Cyber 74. It also has its disadvantages, some of them due to original orientation to cards and batch processing (one can imagine a truly interactive system, more powerful and easier to learn and use). This is an additional reason for an occasional user not to get involved in studying numerous details.

Table 1. Statistical Methods Used by Researchers as Reflected in Theses

Author	Year	Ref.	Work Description	Statistics Used	Comments
Promer, E.D.	68	51	Effect of Spinning Processes on Fiber Bundle Strength & Single Fiber Strength	Mean, Std.dev., C.V.%	Should use T-test
Gunther, D.H.	68	58	Effect of Normal Force Upon Kinetic Coefficient of Friction	Mean, ANOVA	
Simmons, J.F.	68	65	Difference Between Yarn Properties of Yarn Samples Obtained by Shirley Miniature Plant and Full Scale Trial	T-test	Should use F-test too
Whitworth, L.B.	68	47	Effect of Change in Staple Length on Strength and Elongation	Mean, Std.dev.,C.V.%	Should use T-test
Bates, M.R.	70	66	Difference in Strength of Fabrics for Different Yarn Parameters and Different Machines	Mean, Std.dev.,T-test, Progress on Analysis & significance	SPSS could have saved considerable time and efforts.
Cuthburtson, C.R.	70	59	Effect of Fiber Diameter on Friction	Mean, ANOVA	
Ko, F.K.F.	71	61	Relationship of Strength & Elongation with Weaving Parameters	46 Regression Equations & Significance 27 Graphs	SPSS could save time & efforts. Also plots graphs
Asman, L.B.	71	50	Effect of Presence of Dyes in Nylon on Drawing Load	Mean, Std.dev.	Should use T-test
Ogletree, K.K.	71	53	Effect of Bleaching on Fineness, Uniformity & Tenacity	Mean, Std.dev.	Should use T-test
Loti, R.D.	72	49	Effect of Different Surfactants on Fastness	Mean, Std.dev.	Should use T-test
Pulis, R.R.	72	62	Relationship Between Absorbency Coefficient, Scattering Coefficient and Concentration of Dyes	32 Cubic Cur linears equation	Significance of Coefficients as required

Table 1. (cont'd.) Statistical Methods Used by Researchers as Reflected in Theses

Author	Year	Ref.	Work Description	Statistics Used	Comments
Sakai, T.	72	64	Effect of Yarn Strength on Tensile Strength and Distribution	Frequency Tables, Mean, C.V.%, Curvilinears Regression	Significance test on regression equation is essential
Schaefer, G.D.	72		Percent Loops Which Failed to Achieve Desired Heights	Mean, Std.dev.	
Padgett, H.S.	73	60	Change in Stress, Strain and Modulus for Increasing Rate of Strain	Mean, C.V.%, % Change	Significance of difference
Schilling, E.P.	73	67	Comparision of Methods Used for Preventing Static Charge	Mean, Std.dev.	Should use T-test
Mauldin, G.E.	74	55	Effect of Dyeing Processes on Color Difference	ANOVA (repeated 4 times)	SPSS could do the same in one step
Tehrani, H.B.	74	48	Moisture Content After Drying with Microwave Unit	Mean, Std.dev.	
Allen, R.C.	76	54	Difference (in Response to the Questionaire) Between Personal Variable Groups, Numbers, Alumni	Mean, Std.dev., T-test	SPSS has been used for entire analysis
Chao, T.K.T.	76	52	Difference in Strength Between Biaxial & Triaxial Fabrics	Mean, Std.dev. C.V.%	Should use T-test
Knight, L.L.	76	57	Effect of Particle Size, Fabric Weight, Fabric Geometry and Soil Level on Soil Removal	Mean, ANOVA	
Syen, S.	76	63	Relationship Between Energy Consumption and Yarn Properties	Multiple Linear Regression, Curvilinear Regression	Doubtful relationship
Tata, A.S.	76	56	Reproducibility of Dyeing by Observing Color Difference	ANOVA	SPSS could have saved time and efforts.
Coff, H.S.	76	71	Relationship Between Garment Size and Panel Area	Multiple Regression Analysis	SPSS could have saved time and effort

Table 1. (cont'd.) Statistical Methods Used by Researchers as Reflected in Theses

Author	Year	Ref.	Work Description	Statistics Used	Comments
Ko, F.K.F.	76	72	Viscoelastic Response of Textile Fibers	Mean, Curve Fitting	SPSS would help in design of experiment and drawing conclusions.

CHAPTER II

SPSS - COMPUTER PROGRAM PACKAGE

The Statistical Package for Social Sciences (SPSS) is an integrated system of computer programs designed for the statistical analysis of various experimental data. The system was originally developed in 1965 for use in social sciences but can be used equally efficiently in any science or engineering discipline. The package provides a unified and comprehensive system that enables the user to perform many different types of data analysis in a simple and convenient manner.

An SPSS user avoids many difficult, time consuming, and generally unrewarding tasks involved in using a variety of single-purpose computer programs, each with its own idiosyncratic control-card syntaxes and input data formats. The user thus spends less time as a data-preparation clerk and more time as a researcher analyzing results.

The SPSS program package has been implemented and maintained on all the major computer systems (IBM, CDC, DEC, UNIVAC etc.) and is available at many computer installations. The package is well documented [17]. The purpose of this chapter is to introduce the package to the Cyber 74 user in accordance with the aims of this thesis and to build a basis for understanding of the examples of SPSS applications given in the next chapters.

2.1 Setting-up and running an SPSS job

The use of SPSS involves three main components :

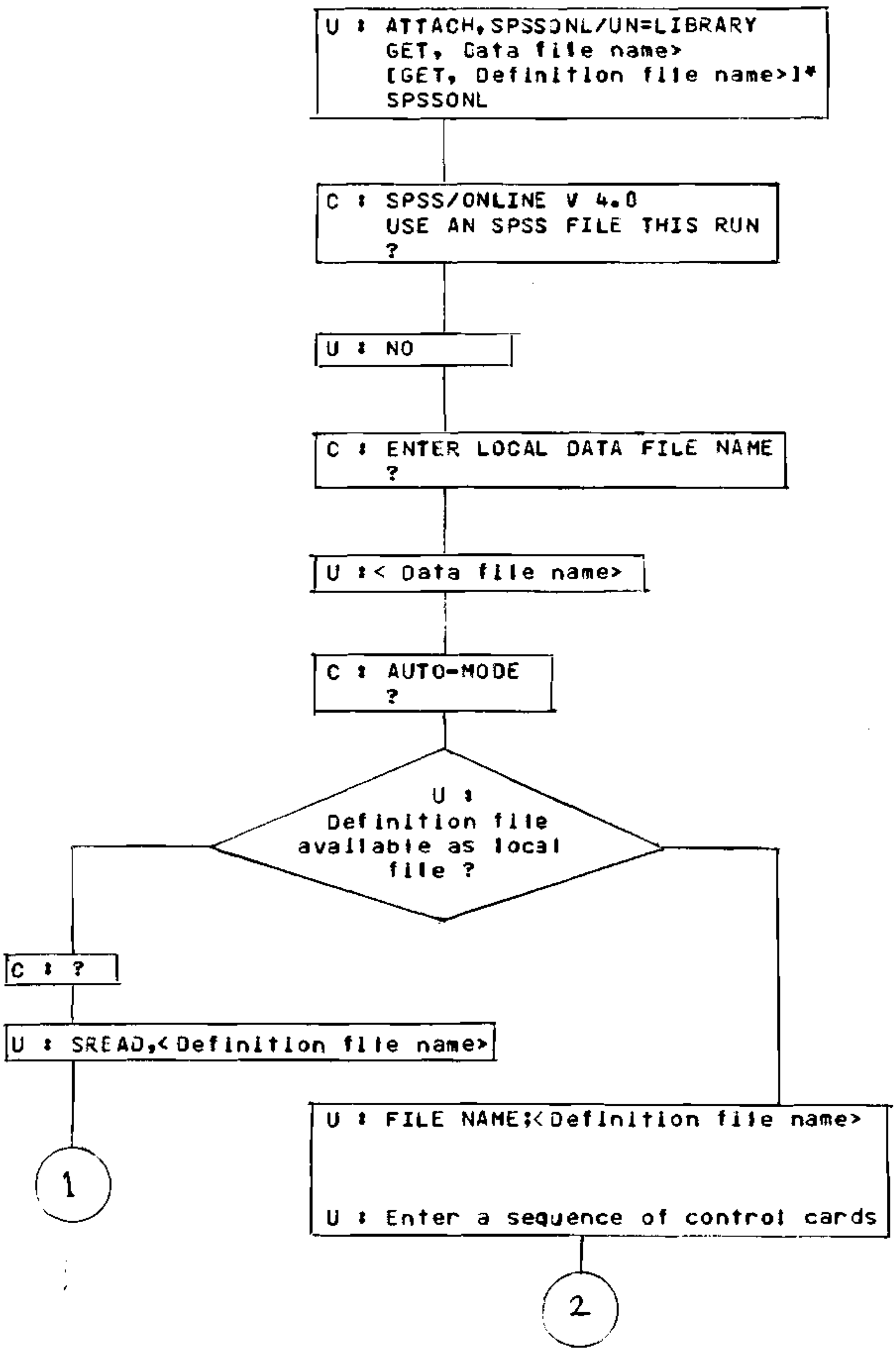
1. A data file - contains user supplied data.

2. An SPSS data definition file - contains data definition cards and procedure name cards, which activate the actual SPSS programs. (the term "card" means an SPSS statement typed on one line).

3. SPSS subprograms - components of the package stored in the computer. The user only need to know how to activate them and is not actually concerned with their detailed structure.

The first step for using SPSS is to make a data file. The method is the same as that for creating a new file on the computer. Once the data file is created, it can be used for any of the SPSS subprograms.

The next step is to create a Definition File. The block diagram of user-computer interaction during running an SPSS job is given in Figure 1. The details are described in the following text.


U : ATTACH,SPSSONL/UN=LIBRARY
GET, <Data file name>
[GET, Definition file name>]*
SPSSONL
C : SPSS/ONLINE V 4.0
USE AN SPSS FILE THIS RUN
?
U : NO
C : ENTER LOCAL DATA FILE NAME
?
U :< Data file name>
C : AUTO-MODE
?
U :
Definition file
available as local
file ?
C : ?
U : SREAD,<Definition file name>
1
U : FILE NAME;<Definition file name>
U : Enter a sequence of control cards
2

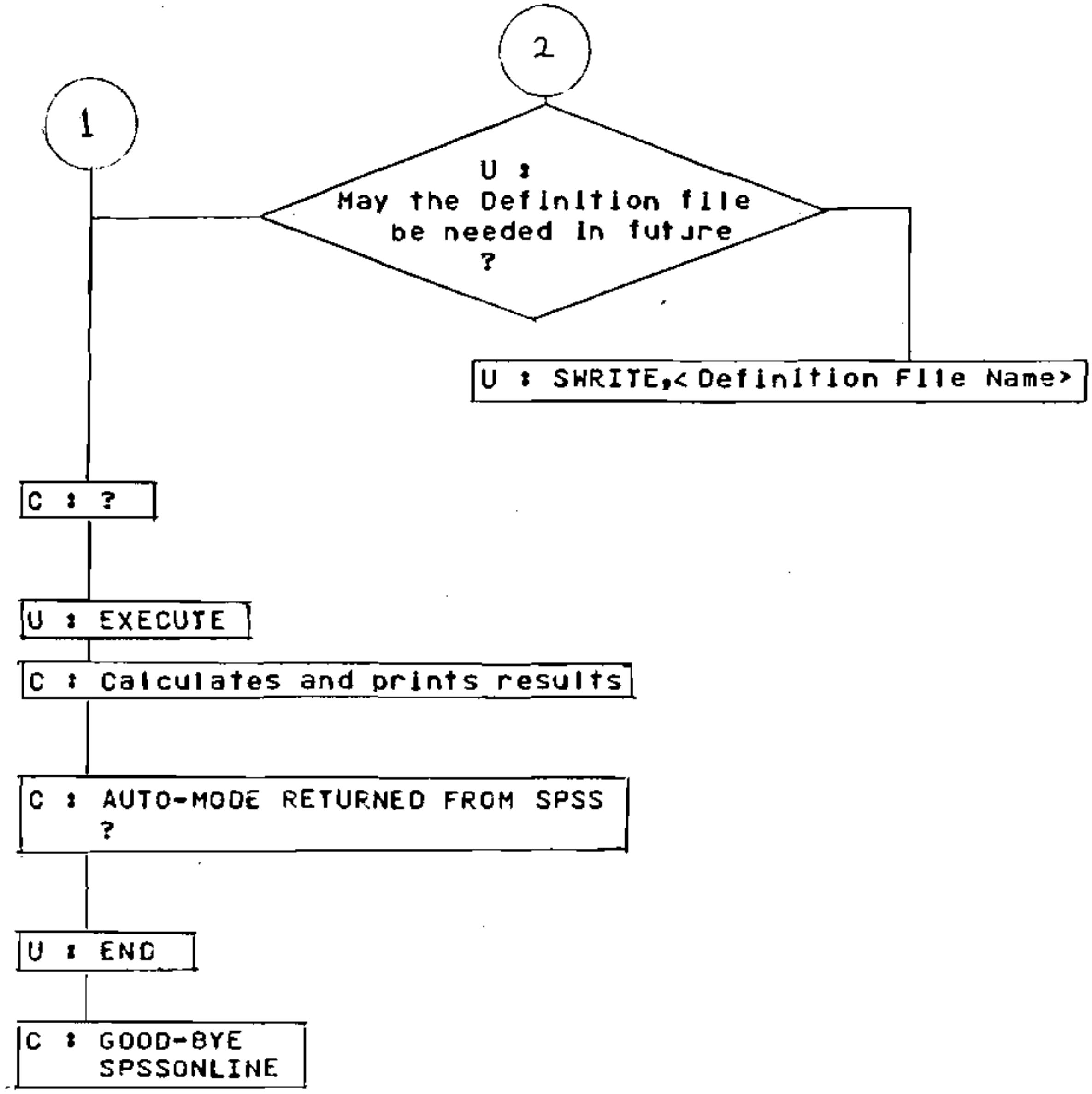


Figure 1. Block-diagram of setting-up and running SPSS job. All capital-letter words are actually printed by a computer (C) or typed-in by a user (U).

2.2 Definition File

Here is a general format of setting-up an SPSS job and creating a Definition File. (the underlined text is entered by the user, the brackets < > are used in place of particular names) :

```
/ ATTACH,SPSSONL/UN=LIBRARY
/ GET, Data File Name>
/ SPSSONL
SPSS/ONLINE V 4.0
USE AN SPSS FILE THIS RUN
?  NO
ENTER LOCAL DATA FILE NAME
?  <Data File Name>
AUTO-MODE
?  5.0 FILE NAME;<Definition file name>
?  10.0 RUN NAME;<A description of the program for heading>
?  15.0 VARIABLE LIST;< Names of variables>
?  20.0 INPUT MEDIUM;DISK
?  25.0 INPUT FORMAT;FREEFIELD
?  30.0 N OF CASES;< Number of cases >
?  35.0 < Procedure name >;< Variable names >
?  40.0 OPTIONS;< Integer number indicating desired option >
```

? 45.0 STATISTICS:<Integer number indicating desired statistics>

? 50.0 READ INPUT DATA

? SWRITE,<Definition file name>

? EXECUTE

The card numbers 5.0, 10.0, 15.0 etc describe the sequence of the procedure cards. Commands are executed in ascending order according to procedure numbers. If the same number is used again in the same file, the previous statement is replaced by the new statement and thus these sequence numbers may be used for editing the file. For continuation of a statement, card numbers like 5.05, 5.10, 10.05, 10.10 etc are used.

Since the above statements are essential for calling SPSS subprograms and are used repeatedly, it would be pertinent to explain each step in detail.

ATTACH,SPSSONL/UN=LIBRARY : This statement is peculiar to the Control Data Corporation Cyber 74 Computers. It attaches the SPSS program from the computer library and without this statement no SPSS procedure card would be valid.

The next step "GET,< file name>" gets a copy of the data file as a local job file from the user's permanent files.

The third statement SPSSONL activates SPSS-Online and the computer responses with a question :

USE AN SPSS SYSTEM FILE THIS RUN ?

At recommended level of sophistication of the SPSS user the answer always must be NO. After the next question

ENTER LOCAL DATA FILE NAME

the user is supposed to type the name of the data file containing the data to be analyzed. When the computer returns with "AUTO-MODE", data definition and procedure name cards are supplied by the user as explained below :

FILE NAME CARD : A file name identifies an SPSS file. This card is optional and can be dropped if so desired by the user along with the RUN NAME card. A file name may be up to eight characters in length and the first character must be alphabetic.

VARIABLE LIST CARD : Each variable in an SPSS file is referred to by a unique variable name. Like all other names used on SPSS control cards, they have a maximum length of eight characters, the first of which must be an alphabetic letter.

The name given to any variable is arbitrary but it is convenient to select the mnemonics which suggest the nature of the variables. For example, if the data file contains strength, elongation and yarn number columnwise, then it would be appropriate to use variable names like STRENGTH, ELONGATE, and YARNUMBR. This would enable to identify the results because all the results are labeled with variable names.

INPUT MEDIUM CARD : This card serves the purpose of informing the SPSS system of the type of input medium from which the user's data will be entered into the system. The control words INPUT MEDIUM are followed by one of the four control words depending upon the location of the user's data (usually DISK).

INPUT MEDIUM;(CARD, TAPE, DISK, OTHER)

N OF CASES CARD : This card simply informs the system of the number of cases in the user's data file. It is very important to remember that the procedure card is "N OF CASES" and not "NO OF CASES" or "NUMBER OF CASES". Any change in the specified procedure name results in an error message and termination of program.

INPUT FORMAT CARD : The first element entered onto the specification field of the INPUT FORMAT card will be one of the two key words, FIXED or FREEFIELD, depending upon whether the user's data is organized in fixed or freefield format.

When the data is in free field format, the system automatically determines the type and location of the variables, because the values of all the alphanumeric variables are set off by single quotation marks and all variable values are seperated by one or more common delimiters.

When the data is in fixed-column format, the keyword FIXED is followed by a FORTRAN or FORTRAN compatible format list giving the type and location of the variables in the data file eg.

INPUT FORMAT;FIXED(2F7.3)

READ INPUT DATA CARD : This card instructs the computer to read from the data file.

SWRITE, < file name> : This card is employed to save the Data Definition File for future reference.

2.3 Rules governing statement order in definition file

The internal order or sequence of most of the statements in a definition file is free with the exception of the following rules :

1. The FILE NAME card, if used, must always be the first.

2. Next is RUN NAME card with a description used for heading.

3. The VARIABLE LIST card follows the above two cards. When these two cards are not used, the VARIABLE LIST card becomes the first in the definition file.

4. INPUT MEDIUM card, INPUT FORMAT card and N OF CASES card follow next in any order after the above cards. The only important rule is that the OPTIONS and STATISTICS card should follow immediately the reference to the corresponding subprogram.

2.4 Summary Of SPSS Procedure Names

Every statistical subprogram has a fixed name in SPSS. No variation in procedure name is tolerated. A summary of procedure names is given below.

1. FREQUENCIES;GENERAL=<Variable name> or ALL: This procedure card calls subprograms for Frequency Distribution, Histogram, Central Tendencies, and Measures of Dispersion.

2. T-TEST GROUPS=<n1,n2>/VARIABLES=<Variable List>: Employed for comparison of two sets of data from the same data file.

3. ONEWAY;Dependent variable BY Independent variable (min,max): Performs Oneway Analysis of Variance.

4. ANOVA;Dependent variable BY Independent variable (min,max): This procedure card is meant for Factorial Analysis of Variance but can as well be used for Oneway Analysis of Variance and would give exactly the same results as obtained by ONEWAY procedure card. The advantage of ONEWAY subprogram is that the CONTRASTS can be evaluated, as explained in Chapter IV.

5. SCATTERGRAM;<Dependent Variable(min,max) > WITH <Independent Variable > : Plots scatter diagram, provides slope, intercept and correlation coefficient.

6. PEARSON CORR;<Dependent Variable> WITH < Independent Variable>: Finds Pearson correlation coefficient and its significance value.

7. REGRESSION; VARIABLES= Variable List>/;REGRESSION = Dependent Variable> WITH Independent Variables>/ :This card can be used for simple regression as well as multiple regression and for cuvilinear regression too.

There are a few other statistical subprograms available in SPSS but the above seven are most frequently used in textiles. SPSS currently does not include all the statistical methods needed for textile purposes. FORTRAN programs have been used in this report where SPSS was inadequate. SPSS cannot perform the Chi-square test for testing any data against hypothetical distributions like the Normal or Poisson, though it covers the Uniform distribution. Other programs not covered by SPSS are analysis of variance for randomized block design, incomplete block design, Latin square and Greco - Latin square.

OPTIONS CARD : Each one of the above subprograms has its own options. The OPTIONS card is employed to control the computation and output.

For example, the OPTIONS card used in conjunction with the FREQUENCIES subprogram enables the user to have frequency tables suppressed and histograms printed. Similarly other subprograms have their own options and the user is advised to refer to SPSS Manual [16, Page Nos.189,242,266,313,314] for the selection of options. If no options or statistics are included, SPSS carries out the computation of the statistics by default.

The OPTIONS card contains the control word OPTIONS followed by a list of integer numbers indicating the options desired within a given subprogram.

STATISTICS CARD : This card is similar in structure and function to the OPTIONS card. This card enables the user to select among a number of available statistics. For example, the subprogram FREQUENCIES allows the user to have any or all of the following statistics computed for the variables in the file; Mean, Standard Deviation, Variance, Skewness, Kurtosis, Minimum, Maximum and Range. Each of these statistics is associated with a certain integer number and may be selected by referring to that number. The user may have all the statistics for a given subprogram reported by typing "STATISTICS;ALL". Of course, the statistics available vary for each subprogram and the user must refer to SPSS Manual for selection of statistics [16, Page Nos. 190,243,246,315].

When all these statistics are computed by the computer as shown in the Illustrations, the computer returns to Auto-Mode for futher instructions.

AUTO-MODE RETURNED FROM SPSS

? END

GOOD-BYE

SPSSONLINE

/SAVE,<Definition file name>

Thus Definition file is saved for future reference. Various set-ups of SPSS jobs will be further exemplified in next chapters and corresponding appendices.

CHAPTER III

DECISION MAKING IN TEXTILE EXPERIMENTATION

3.1 Central tendencies and Dispersion values

Normally the first task of data analysis is to determine the basic distribution characterstics of each of the variables to be used in the subsequent statistical analysis. Information on the distribution, variability and central tendencies of the variables provides the researcher with necessary information required for selection of subsequent statistical techniques and often reveals the basic nature of the data.

For example, the test results of breaking strength of a single yarn are shown in Table 2 and we are interested in the mean and variance of the data.

Mean, variance and skewness could be found manually by using the formulae :

$$\text{Mean} = \bar{X} = \frac{\Sigma X}{N}$$

$$\text{Variance} = s^2 = \frac{\Sigma(X - \bar{X})^2}{(N-1)}$$

$$\text{Standard Deviation} = s = \sqrt{\text{Variance}}$$

$$\text{Standard Error} = \frac{\text{Std. Deviation}}{\sqrt{\text{No. of cases}}}$$

$$\text{\% Coefficient of Variation} = \frac{\text{Std. Dev.} * 100}{\text{Mean}}$$

$$\text{Skewness} = \frac{[(X - X)/s]}{N}$$

$$\text{Kurtosis} = \frac{[(X - X)/s]}{N} - 3$$

Table 2. Results of single yarn breaking strength

NO	STRENGTH	NO	STRENGTH
1	36	11	36
2	37	12	38
3	36	13	33
4	32	14	39
5	36	15	36
6	37	16	39
7	33	17	36
8	34	18	37
9	32	19	32
10	32	20	31

As "higher moments" are attempted, calculation work increases and becomes cumbersome. The above statistics and many more could be obtained by SPSS very easily.

It would be worthwhile to describe the importance of various statistical methods illustrated in this report. Frequently it is desirable to compare two different sets of data with respect to their variability. Since the most common measure of variability is the variance or standard deviation, one might be inclined to compare the standard deviation from each set of data. However,if the data in the two sets represent observation of characterstics which are measured in different units a comparison of standard deviations could be misleading.

For this reason a measure of relative variation is required. Such a measure is the Coefficient of Variation defined as;

$$\%C.V. = \text{standard deviation} * 100 \,/\, \text{Mean}$$

For example, the average single end thread breaking strength of a yarn was 165 grams and the standard deviation was 12 grams. The average elongation noted was 5.2 cm and its standard deviation was 0.5 cm. Which varies more, the strength or elongation?

$$C.V.\%\ \text{of strength} = \frac{12 * 100}{165} = 7.2\ \%$$

$$C.V.\% \text{ of elongation} = \frac{0.5 * 100}{5.2} = 9.6\ \%$$

Though the standard deviation of strength is much higher, there is more variation in elongation than in strength.

The statistical analysis is often based on an assumption of normal distribution of frequencies and it is desirable to know how much the distribution in an analyzed sample differs from normal. The simplest characterstics of deviation of the distribution from normal are skewness and kurtosis.

SKEWNESS : If a distribution is not symmetrical, it is said to be skewed. Skewness in a distribution is caused by a preponderance of extreme values in one tail of the distribution.

For a symmetrical distribution, the skewness is zero. Positive and negative value of skewness indicate that the distribution is skewed on the right and the left sides respectively, as shown in Figure 2.

KURTOSIS : Kurtosis is used to describe the relative peakedness or flatness of the distribution as compared to the normal distribution.

Kurtosis for a normal distribution is equal to 0 and the disrtibution is said to be "mesokurtic". A distribution with

Kurtosis > 0 is more peaked than the normal distribution and is said to be "leptokurtic". If the value is 0 , the distribution is flatter than a normal distribution and is called "platykurtic". Figure 3 depicts the three categories of kurtosis.

Appendix A explains the procedure for obtaining central tendencies, frequency distributions and histograms from the subprograms of SPSS.

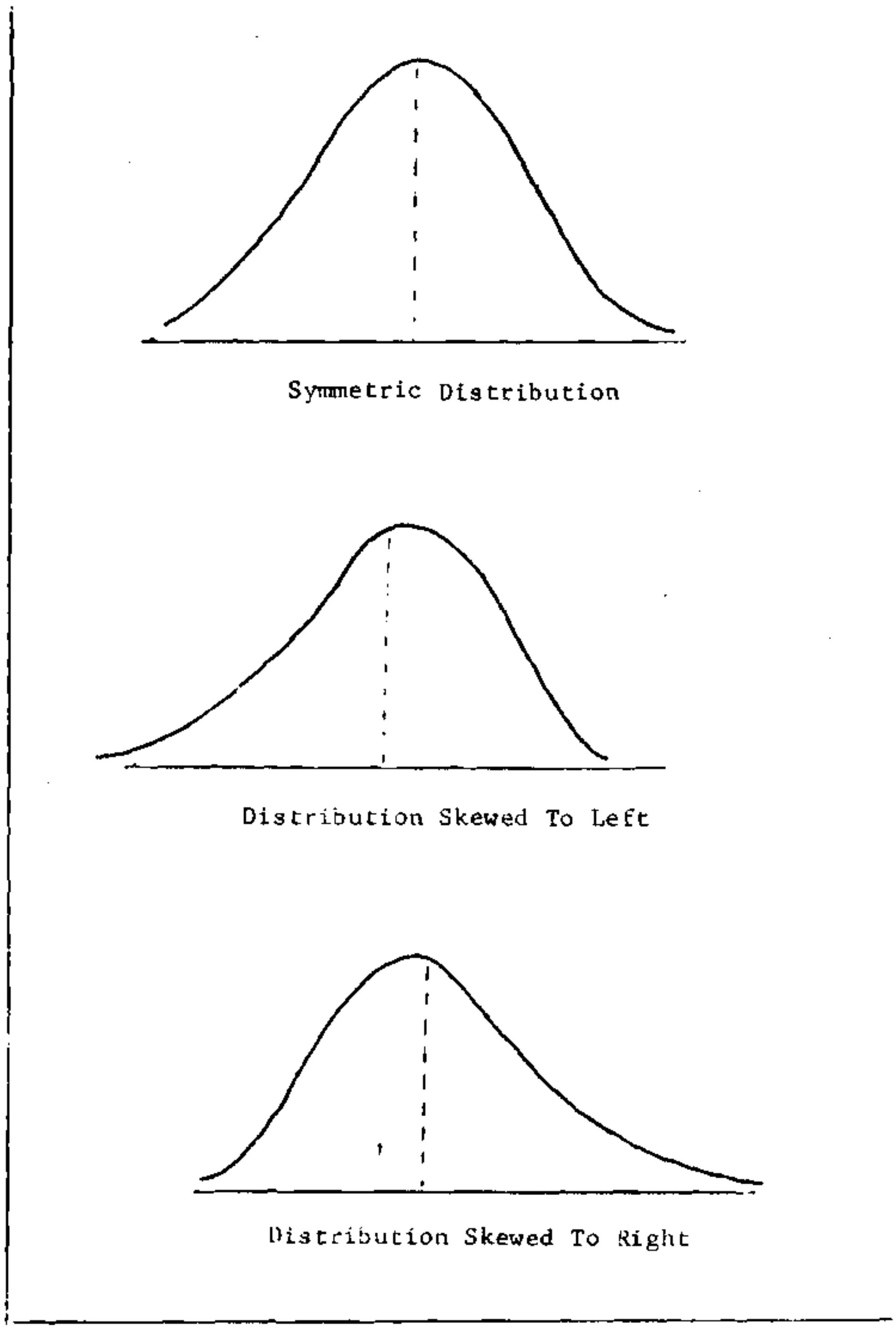


Figure 2. Frequency Polygons Showing Skewness

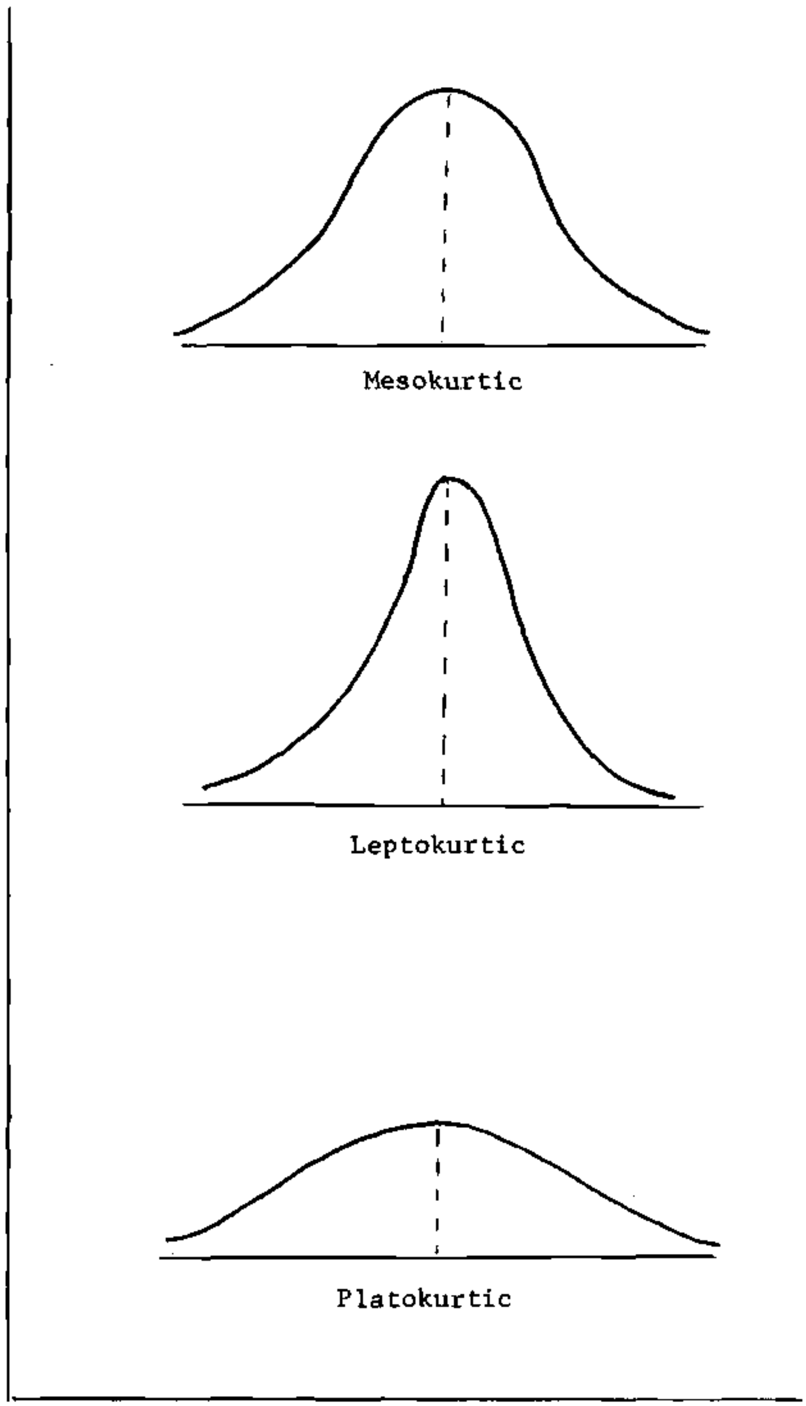


Figure 3. Frequency Polygons Showing Kurtosis

3.2 Frequency distribution tables

The data obtained from experimental results is more or less random as seen in table 1. To get more information and to get it quickly from numerical data, the latter must be organized in some systematic fashion. It is therefore desirable to compress the data into a more compact form and to permit the use of more comlex analytical methods to reveal still other obscure relationships. To condense and simplify data without losing the essential details is the purpose of the analysis of frequency distribution. The tabular presentation of such a summery of data is known as frequency table. The frequency distribution and the SPSS procedure to obtain it, is described in Appendix B. The procedure is similar to the one described in Appendix A for mesures of central tendencies and dispersion values, except that the option 7 supresses the frequency distribution tables.

The advantages of a frequency distribution table are :

1. All the information revealed by the data can be obtained from the frequecy distribution table with greater ease.

2. The frequency distribution not only shows clearly the concentration of individual values but also

brings out the pattern of the tendency for individual values.

3. With the data formed into a frequency distribution, comparisons between two or more series can be made more readily.

4. Frequency tables are indispensible for speeding up computation of other descriptive measures.

3.3 Histogram

It is frequently desirable to draw a picture of a frequency distribution. The most common graphical representation of a frequency distribution is called a histogram.

The frequency in a class is shown by the length of the bar in histograms. Appendix C represents the histogram for the data shown in table 1. The horizontal scale or abscissa, in this case, represents the frequency and vertical scale or ordinate represents the observed measurement.

The advantage of a histogram is that it gives the idea of distribution pattern at a glance. This picture is particularly useful where a large number of measurements have been made.

An interesting point regarding the use of options card may be noted here. The data definition files described in Appendices A, B and C are very similar except that in Appendix A, the selection of option 7 supressed the printing of frequncy tables as stated earlier and in Appendix C, histogram was printed by selecting option 8.

3.4 Comparison Between Two Test Averages

Mills are constantly experimenting with new equipment, new materials and new processing set-ups. From these experiments, conclusions are sought, based upon samples of yarn or fabric tested. Often the new method appears to yield better or worse results than the "regular" method. Yet a difference in sample averages or variation does not by itself prove the superiority of the new method under investigation. The reason for this is that the difference might have been caused by chance fluctuations of sampling, and may therefore not be a true or real difference. Thus the experimenter must perform a statistical test of significance to exclude "chance" as a factor and establish the real nature of the improvement shown in the experiment before recommending to the mill the adoption of the new method.

The statistical evaluation of the mill data brings about the following advantages :

1. Usually fewer tests are needed to come to a decision than by "trial and error" approach.

2. Aids in making better and quicker decision as to whether or not the new experimental equipment, stock or processing set-up represents an improvement in

quality or reduction in production cost over the regular set-up.

Thus statistical evaluation aids the mill in its constant efforts to improve quality and running conditions, reduce seconds and costs.

Evaluation of the significance of a difference in two test averages can best be illustrated from an example in mill processing. A "new" and "regular" method of processing in a mill yielded yarns with single-end elongation to break of 6.4% for the regular method of processing and 7.0% for the new method, under experimental investigation as shown in the Table 3. Does the "new" method represent an actual change in elongation, or the difference is ascribable to chance fluctuations ? T - Test is employed for these types of problems. The statistical procedure for computation of T - Test is as follows :

Mean for first sample = $\bar{X}1$ = 6.4

Mean for second sample = $\bar{X}2$ = 7.0

$$\text{Std. Deviation} = s1 = \sqrt{\frac{\Sigma(X1-\bar{X}1)^2}{N1-1}} = \sqrt{\frac{0.76}{9}} = 0.2906$$

$$\text{Std. Deviation} = s2 = \sqrt{\frac{\Sigma(X2-\bar{X}2)}{N2-1}} = \sqrt{\frac{1.24}{9}} = 0.3712$$

$$\text{Pooled Std. Dev.} = \sqrt{\frac{s1^2\,(N1-1) + s2^2\,(N2-1)}{N1 + N2 - 2}}$$

$$= \sqrt{\frac{0.0844(10-1) + 0.1578(10-1)}{10 + 10 - 2}}$$

$$= 0.3333$$

$$t = \frac{\bar{X}1 - \bar{X}2}{\text{Pooled Std. Dev.}\sqrt{1/N1 + 1/N2}}$$

$$= \frac{7.0 - 6.4}{0.3333\sqrt{1/10 + 1/10}} = 4.02$$

From t- table the value of $t = 2.101$ for 18 degrees of freedom. Since the computed value of $t = 4.02$ exceeds the table value,

the difference in the means of the two samples is significant. Appendix D describes the SPSS procedure to solve these types of problems.

Table 3. Elongation % for regular and new sample.

ELONGATION %-REGULAR			ELONGATION%-NEW		
X1	$X1-\bar{X}1$	$(X1-\bar{X}1)^2$	X2	$X2-\bar{X}2$	$(X2-X2)^2$
6.4	0	0	7.2	0.2	0.04
6.5	0.1	0.01	6.8	-0.2	0.04
6.3	-0.1	0.01	6.5	-0.5	0.25
6.1	-0.3	0.09	6.7	-0.3	0.09
6.9	0.5	0.25	6.8	-0.2	0.04
6.3	-0.1	0.01	6.7	-0.3	0.09
6.9	0.5	0.25	7.2	0.2	0.04
6.2	-0.2	0.04	7.4	0.4	0.16
6.1	-0.3	0.09	7.0	0.0	0.0
6.3	-0.1	0.01	7.7	0.7	0.49
---	----	----	---	----	----
64.0		0.76	70.0		1.24

CHAPTER IV

TESTING DIFFERENCES IN RESULTS OF EXPERIMENTS

(Analysis of variance)

Progressive management expects, their technologists to suggest new methods of improving the quality of manufactured poduct, lowering production costs, and increasing productivity. Improvements in processing involve changes in methods and equipment. Yet the changes can not be brought about on the basis of logic and common sense alone. Facts and data must support and justify the changes to be made. These facts and figures come from pilot-plant experiments followed by full-scale mill trials

The planning of experiments is the most critical phase of the experimental program, and the theory of statistics is vitally concerned with the framework in which the experimenter can state his questions. An application of the general principle of the Design of Experiments can be explained by a practical example.

4.1 Oneway Analysis Of Variance

In a Textile mill four different speeds for ring frames were tested and breakage rate per 10000 spindle hours was noted which appears in Table 4.

Table 4. Breaks/10000 spindle hours for different speeds.

SPEED=1	SPEED=2	SPEED=3	SPEED=4
56	64	45	42
55	61	46	39
62	50	45	45
59	55	39	43
60	56	43	41
---	---	---	---
292	286	218	210

Do these speeds effect the breakage rate? Let our null hypothesis, H0, be that there is no difference in breakage rate with different speeds and let the alternative hypothesis H1, be that there is a differene in breakage rate with different speeds. The method of analysis of variance for testing these hypotheses is as follows :

$$\text{Correction Factor} = C = \frac{(\sum \sum Xij)^2}{N} = \frac{(1006)^2}{20} = 50601$$

$$\begin{aligned}\text{Total sum of squares} &= \sum \sum Xij^2 - C \\ &= 56^2 + 55^2 + \ldots\ldots\ldots + 43^2 + 41^2 - 50601 \\ &= 1338.2\end{aligned}$$

$$\begin{aligned}\text{Treatment sum of squres} &= \frac{\sum(\sum Xij)^2}{r} - C \\ &= \frac{292^2 + 286^2 + 218^2 + 210^2}{5} - 50601 \\ &= 1135.0\end{aligned}$$

SOURCE	S.S.	D.F.	M.S.	F-EVAL	F-TAB
Treatments	1135	3	378.3	29.79	3.24
Error	203.2	16	12.7		
Total	1338.2				

(S.S. stands for sum of squares, D.F. means degrees of freedom, M.S. is mean sum of sqares. F-EVAL represents calcuated value of F and F-TAB is the F value from statistical table for corresponding degree of freedom. These notations have been used throughout the text.)

A difference may be noted in the F value from statistical table shown in the text and F probability (or significanc) values in the Appendices. The F table value in the text represents the F value for corresponding degrees of freedom at 95% confidence level. This value is compared to the calculated value of F to detect the significance of a hypothesis.

But the F probability value in Appendices is the value (type I error). If the hypothesis is really true and is rejected by the sample, type I error is committed. A probability value of 0.05 or less means that the probability of making type I error is 5 % or less, which is an acceptable risk in quality control.

Error sum of squares is the difference between the total sum of squares and the treatment sum of squares. Since the calculated F-value (29.79) exceeds F - probability value (3.24) found in statistical tables, we accept the alternate hypothesis that there is a difference in breakage rate with different speeds.

Now the problem arises : Which of these four speeds have significant effect on breakage rate ? This can be determined by using "Contrasts". Let, for instance, C1, C2, C3 be the Contrasts between SPEED-1 and SPEED-4, 2 and 3, and average of 1+4 and 2+3, respectively.

Contrasts	T1	T2	T3	T4
C1	+1	0	0	-1
C2	0	+1	-1	0
C3	+1	-1	-1	+1

$$C1 = 1(292) + 0(286) + 0(218) - 1(210) = 82$$

$$C2 = 0(292) + 1(286) - 1(218) + 0(210) = 68$$

$$C3 = 1(292) - 1(286) - 1(218) + 1(210) = 0.2$$

$$\text{Sum of the squares of Contrast 1} = \frac{(82)^2}{5*2} = 672.4$$

$$\text{Sum of the squares of Contrast 2} = \frac{(68)^2}{5*2} = 462.4$$

$$\text{Sum of the squares of Contrast 3} = \frac{(-2)^2}{5*4} = 0.2$$

Each of the contrasts have 1 degree freedom.

SOURCE	S.S.	D.F.	M.S.	F-EVAL	F-TAB
Contrast 1	672.4	1	672.4	52.9	4.43
Contrast 2	462.4	1	462.4	36.4	4.43
Contrast 3	0.2	1	0.2	0.1	4.43
TOTAL	1135	3	378.3		
Error	203.2	16	12.7		

Since in the first two cases computed F-value exceeds Table F-value, it can be concluded that treatments 1 and 4, as well as 2 and 4 are different. Appendix E explains how all this computation can be done easily by SPSS.

4.2 2 - Factorial Experiment

In the previous example, the effect of one treatment was studied. But in factorial experiment, there are two or more independent classes of interest. Moreover, here not only the effects of independent treatments are studied but also their interaction is taken into considerarion.

To explain interaction, let us consider an experiment conducted to determine the durability of rings of spinning frames and the factors of interest are three different kinds of travellers and four different spindle speeds. In such a case, a factorial experiment not only gives the effect of the type of travellers and spindle speeds, but also their possible joint effect is determined. This differential effect is known as "interaction".

Some advantages of a factorial experiment can be seen

1. They are more efficient than one factor-at-a-time experiments.

2. The same data are used in investigation of the effect of both factors.

3. Some informatiion is gained on possible interaction between the two factors.

EXAMPLE : Fabric samples were dyed with three types of dye-stuffs and with two concentrations for each dye. Dye penetration was tested for these two variables. The results were as shown in the Table 5.

Table 5 - Dye penetration with different dyestuffs

and concentrations.

	Type of Dyestuff			
Concn.	A	B	C	
	9	14	1	
0.5	12	15	6	(93)
	13	16	7	
	--	--	--	
	34	45	14	
	4	3	2	
1.0 %	8	5	10	(78)
	18	17	11	
	---	---	---	
	30	25	23	
	64	70	37	

The statistical procedure for the factorial analysis of variance is explained below :

$$\text{Correction Factor} = \frac{(\Sigma\,\Sigma\, Xij)^2}{N} = \frac{171^2}{18} = 1624.5$$

$$\text{Total sum of squares} = \Sigma\, Xij^2 - C$$

$$= 9^2 + 12^2 + \dots\dots\dots + 11^2 - 1624.5 = 484.5$$

$$\text{Dyestuff sum of sq.} = \frac{\Sigma(\Sigma\, Xij)^2}{r} - C$$

$$= \frac{64^2 + 70^2 + 37^2}{6} - 1624.5 = 103$$

$$\text{Conc. sum of sq.} = \frac{\Sigma(\Sigma\, Xij)^2}{n} - C$$

$$= \frac{(93^2 + 78^2)}{9} - 1624.5 = 12.5$$

$$\text{Sum of squares of subgroups} = \frac{(34^2 + 45^2 + \dots\dots + 23^2)}{3} - 1624.5$$

= 185.8

Conc & Dyestuff interaction = 185 - 103 - 12.5

= 70.3

SOURCE	S.S.	DF	M.S.	F-EVAL	F-TAB
Dyestuff	103	2	51.5	2.07	3.89
Conc.	12.5	1	12.5	0.5	4.75
Conc * Dyestuff	70.3	2	35.1	1.49	3.89
Error	298.7	12	24.9		
Total	484.5	17			

Since the calculated values of F are less than the F values from the statistical tables for the corresponding degrees of freedom, it could be concluded that neither concentration nor the type of dyestuff significantly affect dye penetration.

SPSS program for this example is shown in Appendix F. It should be noted that there is a slight difference in making a data file for data having multiple classification and that for single factor analysis of variance. The first column of the data file in Appendix F represents Concentration % and since there are two concentrations tested, the first column has a value of either 1 or 2. Similarly the second column of the data file indicates type of Dyestuff and because there are three kinds of dyestuffs in the experiment, column 2 takes values of 1, 2 or 3. The third column represents the dye penetration. The variables in the SPSS file have been defined in the same sequence. The order of variables is not important. The only point is that the data need to be defined in the SPSS file in the same sequence.

4.3 2 - Factorial Experiment

The third order factorial experiment is just a step further to the second order experiment illustrated in the previous example. The procedure remains the same. Here the effect of three independent variables is studied on a single dependent variable. So in this case one more variable is added to the data file and data defining SPSS file.

EXAMPLE : The data file in Appendix F contains three independent variables viz. cotton mixing, feed in combing and number of needles. The dependent variable is neppiness in the combed sliver. It is obvious from the data that two mixings have been tried with three feed rates and five different numbers of needles. SPSS data defining file defines these variables in the same order. The program and results are self-explanatory.

CHAPTER V

TESTING DIFFERENCES IN FREQUENCY DISTRIBUTION

(Chi-square test)

Variation in end breakage among spindles on a spinning frame is by and large random in nature. The number of spindles having a certain number of breaks tends to follow a Poisson Distribution, provided no systematic sources of variation are present.

This means that purely as a matter of chance a large number of spindles will have no end breaks, many spindles will have only one or two, and a few spindles will have a large number.

A simple procedure for determining if spindles are in control or out of line is to record the number of breaks that occur during a spinning test and write the frequency distribution of these spindles having a varying number of breaks.

This actual distribution can then be compared with the one that could be expected if it were to follow a Poisson distribution. Whether the difference between the actual and expected distribution is significant can be determined by using the CHI- square formula

CHI-SQ = [(Actual Frq - Exp Frq)**2] / Exp Frq where Actual Frq is the number of breaks observed during the experiment and the Exp Frq is the expected frequency of breaks as per Poisson distribution.

This calculated chi-square value is compared with the chi-square table value for (k-2) degrees of freedom where k is the number of classes in the experiment. If the calculated value exceeds table value, it can be concluded that the distribution did not follow Poisson distribution. In other words, the breaks are not due to chance but some systematic source of variation is present. The observed number of breaks and the method of calculating the chi-square value is shown in Table 6 and explained further immediately after the table.

We wish to determine now how the observed frequencies would compare with the expected frequencies of a Poisson distribution having the same mean and the same total number of frequencies.

Table 6. Frequency of breaks and Poisson distribution.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	61	0	0.1827	48.23	12.77	3.38
1	92	92	0.3160	82.00	10.00	1.22
2	53	106	0.2640	69.70	-16.7	4.00
3	29	87	0.1496	39.49	-10.49	2.79
4	12	48	0.0636	16.79	-4.79	1.37
5	8	40	0.0216	5.70		
6	3	18	0.0061	1.61		
7	3	21	0.0015	0.40	9.2	10.82
8	-	-	0.0003	0.08		
9	1	9	0.0001	0.03		
11	1	11	0.00	0.00		
16	1	16	0.00	0.00		
	---	---				---
	264	448				23.47

Mean = 448/264 = 1.70

Column 1 represents the number of breaks and column 2 is the frequency of occurrence of the corresponding number of breaks. Column 3 denotes the total number of breaks in each class and is the product of column 1 and 2. In Column 4, the probability distribution for Poisson having the same mean i.e. 1.70 is posted. This information is obtained from Poisson distribution Function Table for mean = 1.70 and x = breaks per spindle. Column 5 is the expected number of breaks and is the product of column 2 and 4. In column 6, the difference of observed and expected frequency is computed.

The chi-square test holds good only if frequency in each class is at least 5. But it can be observed from Column 5 that the last 6 rows have a smaller value than 5. Hence they are all added up to the 6th row. The number of classes is now reduced to 6 and so the degree of freedom (k-2) is taken as (6 - 2) = 4 instead of the previous number of classes k=12.

Column 7 is the square of the difference of observed and expected frequency divided by the expected frequency, the sum of which is Chi-square value.

Chi-square value from Table = 9.49 for 4 d.f.

Since the computed value exceeds the table value, there is a reason to believe that the breaks are out of control. Appendix H describes the FORTRAN Program for testing hypothesis for Poisson distribution. It finds the value of

Chi-square from the data fed by the user which only needs to be compared with the Table value of Chi-square for k-2 degrees of freedom, where k is the number of classes in the actual distribution.

CHAPTER VI

PRACTICAL APPLICATIONS OF REGRESSION ANALYSIS

Regression analysis is the study of the change in one variable associated with the change in another. The relationship expressed mathematically is commonly known as Regression Equation. A simple regression is used to estimate the value of a dependent variable (Y) for a given value of an independent variable (X). A multiple regression equation incorporates more than one independent variable. In textiles, the relationship between fiber properties and yarn strength or processing efficiency are often expressed in terms of simple or multiple regression equations. The precision of the estimated values obtained from a regression equation is measured by the standard error of estimate.

A regression equation, especially a multiple regression equation, can provide a wealth of information on the relationship between dependent and independent variables. For example, yarn strength can be predicted from regression equation relating fiber length, fiber fineness, and Pressley Index. Cotton buyers and production managers frequently substitute one fiber property for another because of the supply and demand situation of certain cotton qualities in the market or in the warehouse. So the defficiency of one property can be compensated for by improving another property maintaining constant yarn strength.

6.1 Scatter Diagram

Before any steps are taken towards developing a functional relationship it is necessary to consider what form this function should take. The logical procedure would be to plot the data on what is known as a Scatter Diagram. Each observation is a point on the graph with the usual convention that the independent variable is plotted along the abscissa(X-axis) and the dependent variable on the ordinate (Y-axis). The pattern of these points should suggest a functional form which is most suitable to fit the data. It is expected that the points will not fall along a single curve and a certain amount of variation can be expected.

EXAMPLE : Thermal Resistance of a carpet was studied for different pile heights. The results are given in Table 7.

Table 7. Thermal resistance and pile height of carpet samples.

S N	R-val(Y)	Ht.(X)	S N	R-val(Y)	Ht.(X)
1	0.68	0.19	10	0.95	0.37
2	0.65	0.18	11	1.66	0.65
3	0.67	0.23	12	1.96	0.70
4	0.55	0.15	13	2.46	1.06
5	1.12	0.38	14	2.19	0.60
6	1.33	0.48	15	1.83	0.77
7	1.51	0.39	16	1.90	0.67
8	0.78	0.21	17	1.71	0.63
9	1.03	0.21	18	0.70	0.14

Appendix I describes how to obtain Scatter Diagram and Correlation Coefficient by SPSS. The statistical procedure for calculating correlation coefficient is given below.

In the present problem :

$\Sigma X = 8.01$ $\Sigma X^2 = 4.7503$ $\Sigma XY = 13.0973$

$\Sigma Y = 23.68$ $\Sigma Y^2 = 37.3294$

Correlation coefficient = r

$$= \frac{n\Sigma XY - \Sigma X \Sigma Y}{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}$$

$$= \frac{18*13.10 - 8.01*23.68}{(18*4.75 - 64)(18*37.33 - 544.29)}$$

$$= 0.9458$$

6.2 Linear regression

From the correlation coefficient of the last example, it was quite clear that thermal resistance and pile height of the carpet are strongly related. However, linear regression can predict the change in thermal resistance when the pile height is changed.

The SPSS procedure cards for linear regression appear in Appendix J.

Assuming the relationship $Y = a + b X$, where b is the coefficient of regression and a is Y intercept, the statistical procedure for computation of this equation is as follows :

$$b = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{18*13.1 - 8.01*23.68}{18*4.75 - (8.01)}$$

$$= 2.159$$

$$a = \frac{Y}{n} - \frac{b*X}{n}$$

$$= \frac{23.68}{18} - \frac{2.159*8.01}{18}$$

Here again it would be appropriate to describe some of the important terms used in regression analysis.

REGRESSIION ANALYSIS It attempts to establish the "nature of relationship" between variables so that we may be able to predict the value of one on the basis of another.

$$Y = a + b X$$

Conventionally, the variables which are the basis of prediction are called "independent variables" and shown as X. The predicted variable is called "dependent variable" and shown as Y.

REGRESSION COEFFICIENT It is the slope "b" of the line of regression with reference to X-axis.

CORRELATION ANALYSIS : It determines the degree of agreement between the variables. Here depedent and independent variable is a personal choice and is of no practical significance.

STANDARD ERROR OF REGRESSION The standard deviation about the line of regression is called the standard error of regression.

$$Sxy = \frac{Y- a \quad Y - b \quad XY}{n-2}$$

A standard error of zero will result when all the points fall on the regression line, indicating a perfect fit of the sample data. It enables us to evaluate the dispersion about the line of regression.

COEFFICIENT OF CORRELATION It is the ratio of explained variation to the total variation. Hence the coefficient r determines the degree to which the regression equation explains the variables.

$$r = \frac{n\sum XY - \sum X \sum Y}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$r = (\text{sign } b)\sqrt{r}$ taking the sign of b, the slope.

The value of correlation coefficient 0 indicates no relationship whereas 1 denotes perfect relationship. The value r with plus sign indicates a direct relationship and a minus sign an inverse relationship.

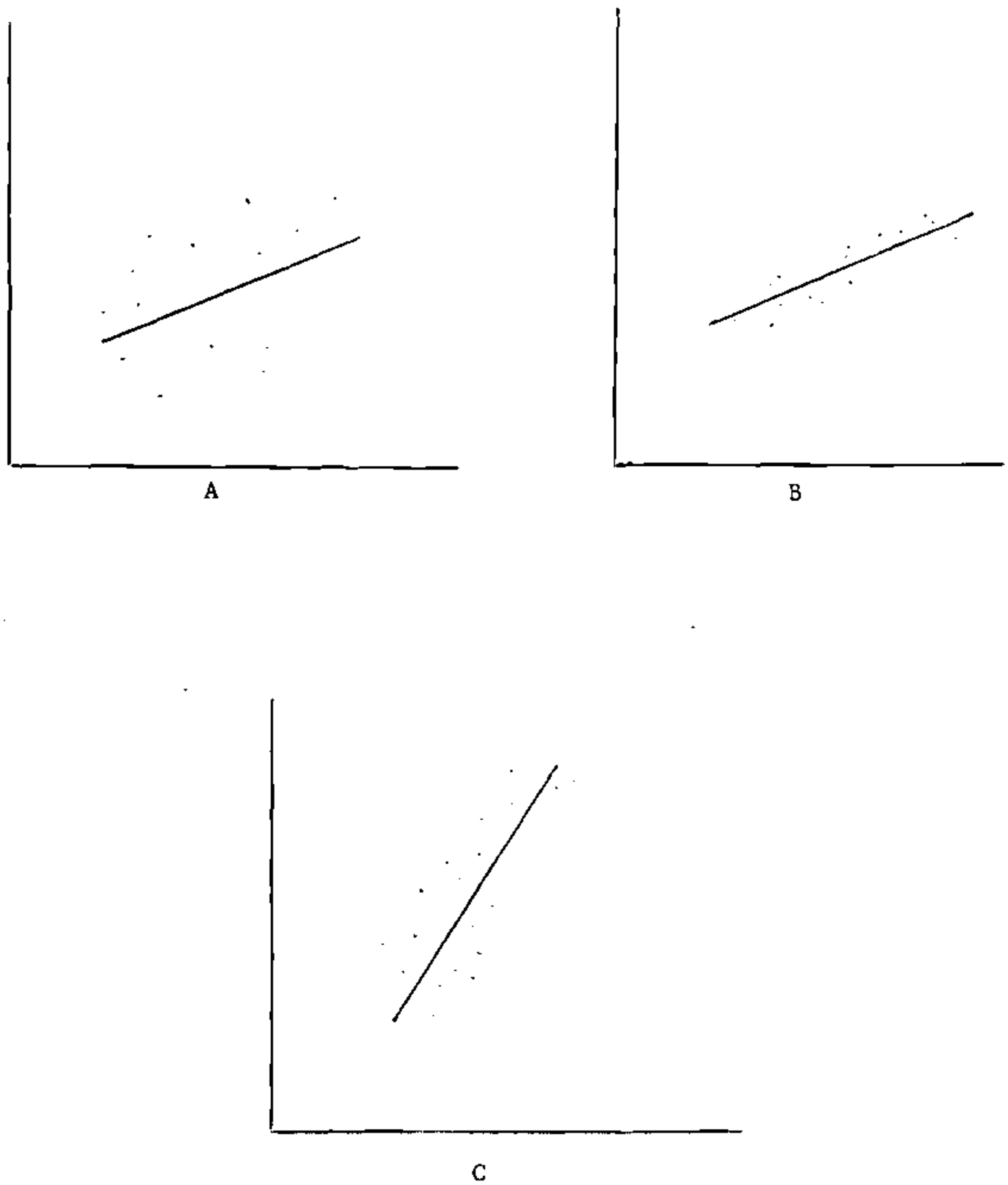


Figure 4. Comparison of Three Scatter Diagrams

The three measures viz, Regression Cofficient or slope b, Standard Error of Regression Sxy and Correlation coefficient have different purposes as evident from Figure 4. C has highest regression coefficient since slope is highest. B has lowest standard error of regression as dispersion is lowest in this case.

For Correlation coefficient r, no conclusion can be drawn from these figures and calculation is necessary, because it is the ratio of explained and total variation and would depend on numerator as well as denominator.

6.3 Multiple regression and correlation

As stated earlier, the multiple regression and correlation analysis involves the study of association among more than two variables. The subject is a logical extension of simple regression and correlation analysis with the same interest in predictive relationship and dispersion values and measures of the degree of association. The basic theory underlying the techniques of analysis also parallels the development of simple regression and correlation, and the SPSS procedure is exactly same except that more variables are added to the variable list and regression procedure card.

When a regression study involves more than two variables, there is still a single dependent variable which we would like to predict, but now there are two or more independent or predicting variables. The symbol X1 is used for the dependent variable and X2, X3, etc for several independent variables. This system provides a more convenient subscripting procedure. The words "independent variables" imply that they are independent, each from the others. In actual studies it is quite possible that there may be a certain amount of interdependence among so-called independent variables which may influence the interpretation of results. In many cases, several independent variables which may be intially included in the study with the expectation that some less important variables may be dropped at a later time.

Table 8. Marker Bank

```
------------------------------
DUA5900201,DUA,59,2.50,89.15
3432,1
3632,1
0,00
DUA5900202,DUA,59,2.32,88.81
3329,1
3627,1

   .

   .

   .

DUA59010001,DUA,59,12.80,88.23
3030,1
3630,1
3730,1
3828,1
3829,2
3830,1
3929,1
3931,1
4033,1
0,00

   .

   .

   .

DUA5901313,DUA,59,17.04,88.49
3232,1
3428,1
3529,1
3628,1
3729,1
3828,1
3930,1
4029,1
4131,1
4228,1
4332,1
4431,1
4632,1
0,-1
------------------------------
```

Let us consider the following example : Men's trousers are marked for cutting with different distribution of sizes in markers (or panel layouts). There is a bank of over 100 markers accumulated over a period of time. The description of the markers as shown in Table 8 and stored in the computer in the following format :

Marker ID, Style ID, Fabric width, Marker length, Utilization%

Size, Quantity

Size, Quantity

Size, Quantity

.

.

.

.

There is a large variety of sizes designated by four-digit number with the first and the last two digits standing for in-seam length and waist respectively. Approximate values of the panel areas for each single size were required. It was impossible to find their values directly from the marker description because each marker represented a mixture of sizes. However, it was expected that in general a panel area A is a linear function of in-seam length L and waist W :

$$A = a + b \cdot L + c \cdot W$$

or in standard notation

$$X1 = a + b2 . X2 + b3 . X3$$

It was possible to find the coefficients a, b, c (or a, b2, b3) by using each marker as a point in three-dimensional space X1, X2, X3,

where X1 = Panel area

X2 = Sum of length in each marker

X3 = Sum of waists

and by applying a standard multiple regression technique; the latter offers the following normal equations for evaluation of regression coefficients in a case of two independent variables :

The normal equations for two independent variable cases are

$$\Sigma X1 = na + b2 * \Sigma X2 + b3 * \Sigma X3$$

$$\Sigma X1X2 = a * X2 + b2 * X2 ** 2 + b3 * X2X3$$

$$\Sigma X2X3 = a * \Sigma X3 + b2 * \Sigma X2X3 + b3 * \Sigma X3 ** 2$$

where a, b2, b3 are the coefficients of regression equation.

Solving these equations simultaneously gives the values of the regression coefficients a, b2, b3. It is also possible to estimate the closeness of the experimental points to the regression plane, the standard error and the confidence intervals for regression coefficients.

Coefficient of Multiple Determination and Multiple Correlation Coefficient :

$$R\ 1.23 = \frac{a\Sigma X1 + b2\Sigma X1X2 + b3\Sigma X1X3 - (\Sigma X1)^2/n}{X1^2 - (\Sigma X1)^2/n}$$

= 0.999 (from Appendix K)

Coefficient of Multiple Determination is the ratio of explained variation and total variation. The value R-sq = 0.999 means that 99.9 % of variation in X1 has been explained by the variables X2 and X3. included in this data.

R is called Multiple Correlation Coefficient. The value of Multiple Coefficient of Correlation ranges from 0 for no relationship to 1 for a perfect relationship. It indicates the collective degree of association between the several independent variables and the dependent variable.

Appendix K shows how to use SPSS for multiple regression analysis. The experimental data was rather voluminuous in our example (the Table 7 shows only a small part of the marker bank). Therefore instead of manual calculation and insertion of X1, X2, X3 simple FORTRAN program was used for reading from the marker bank (on Tape 7) and creating automatically an SPSS data file on Tape 8.

In order to secure better precision of the results values of area were multiplied by 100. The results show, after scaling the coefficients down to actual values, the panel area (in sq. yards).

$$A = 0.02979\ L + 0.02502\ W - 0.00006$$

The significance value of .000 in Appendix K means that our hypothesis about linear dependency of panel area on length L and waist W may be taken for true with high degree of confidence. The exception is the significance value .999 for the free coefficient -0.00006, presumably due to its proximity to zero. It seems obvious that this coefficient may be dropped from the equation altogether.

6.4 Multiple Regression and Fabric Geometry

The paper by Konopasek and Wustrow [10] on computation in engineering design of knitted fabrics referred to the following experiment :

Plain jersey fabric was produced from cotton/poyester yarn of 18, 20 and 24 cotton count knitted on machines with 16, 18 and 20 cut, with very loose, normal and very tight settings (i.e. altogether 27 different fabrics). Measurements were taken for yarn length per revolution, course and wale density, fabric width and yeild in grey and finished fabric, repeat height, length and width shrinkage in washtest, bursting test. An existance of a relationship between the coefficients K, KC, KW and shrinkage SC and SW was suspected. The following notations were used.

K = yarn linear density / stitch length.

KC = stitch length / course spacing ratio

KW = stitch length / wale spacing ratio

SC = length shrinkage

SW = width shrinkage

Multiple regression analysis was employed to find out what degree of the width shrinkage SW and the length of shrinkage SC were responsible for changes in KW and KC.

The data, SPSS procedure and results appear in Appendix L. The multiple regression equation resulted as:

$$KC = -0.333 + 0.312\ K + (0.029\ SC) + 0.400\ SW$$

$$KW = 6.500 - 0.198\ K + 0.129\ SC + (0.020\ SW)$$

SPSS procedure not only predicts the relationship but also its significance. The components in brackets were found statistically insignificant bacause the standard error of these two cases were higher than the coefficients as can be seen in the Appendix L. What is left shows that an exessive width stretch of the fabric during calendering does not affect course-wise shrinkage and reduces wale-wise shrinkage. Similarly excessive length stretch does not affect wale-wise shrinkage and reduces course-wise shrinkage. This negative impact may be explained as a consequence of resin treatment which makes the fabric more stable.

Thus multiple regression can be conveniently used for determining significant relationship between several parameters and also their significance can be evaluated.

Curvilinear Curve Fitting

In many situations a linear function will not suit the relationship between the two variables. This may be obvious from the pattern shown on the scatter diagram or from an advanced knowledge of the theoretical relationship. Under these circumstances it is frequently possible to transform either one or both of the variables to a linear form to accommodate the patterns of the scatter diagram. Another possibility is to use the polynomial function of the lowest order necessary to obtain a good fit. Let us first consider variable transformation.

Among the simpler variable trasformations are the reciprocals, exponential, and power functions. These functions are of the form shown in the following equations and diagrams

Reciprocal function Y = a + b / x

Exponetial function Y = a * b ** X

Power function Y = a * X ** b

Notations like "*" for multiplication, "**" for exponential and "/" for division have been used in the text for convenience. Some typical nonlinear relations are shown in figures 5, 6 and 7.

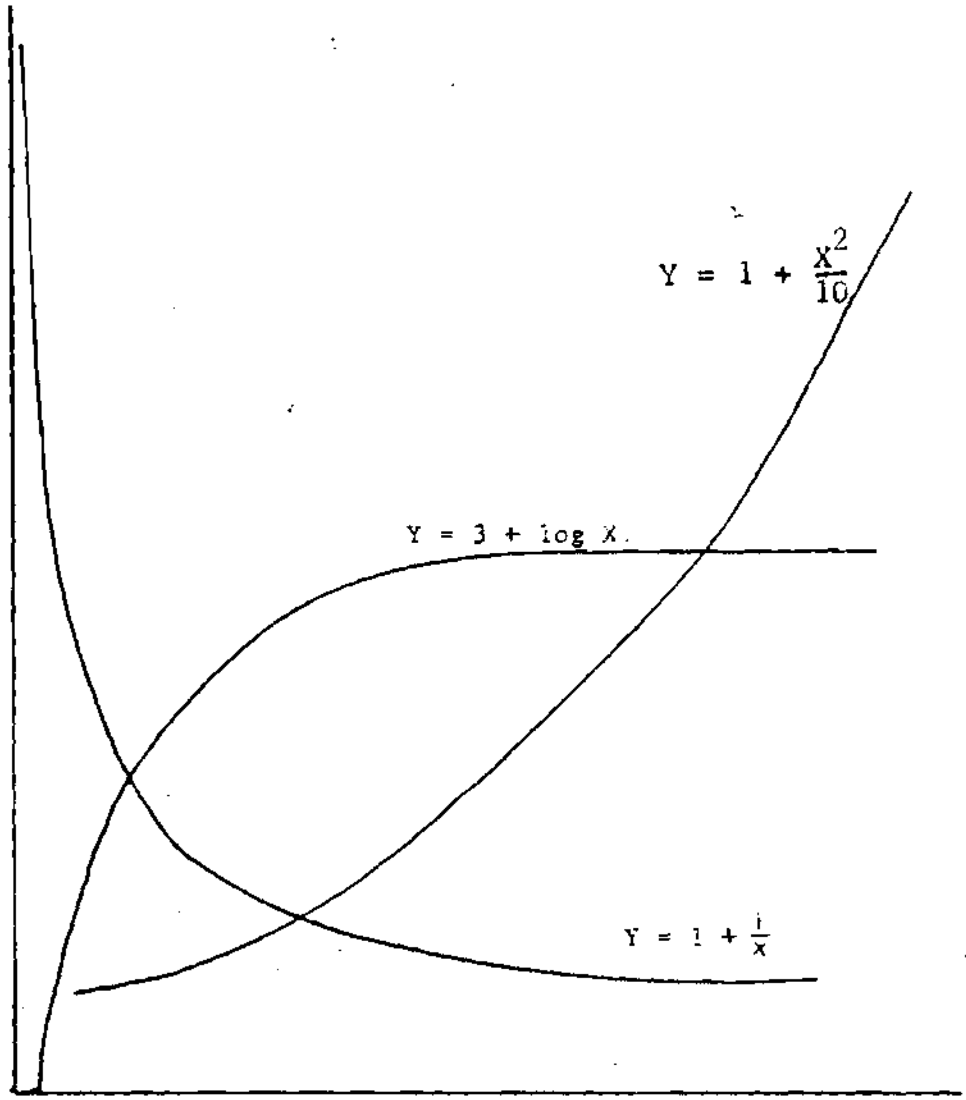
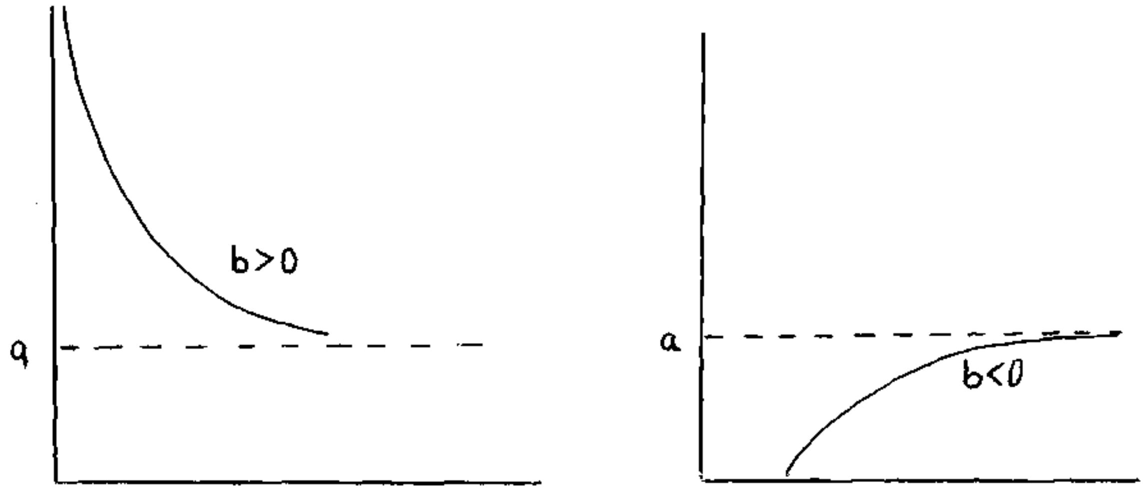


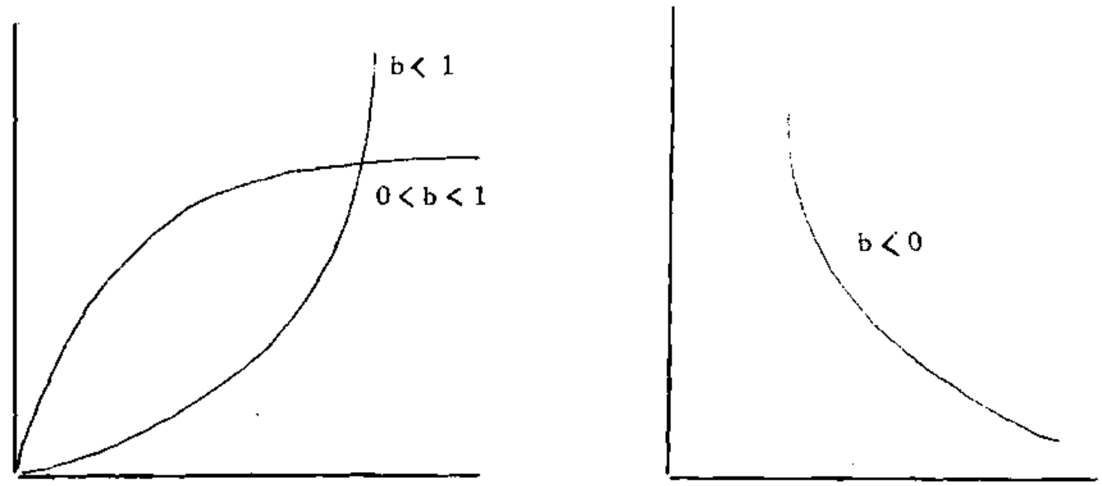
Figure 5. Curvilinear Regression Equations And Their Shape



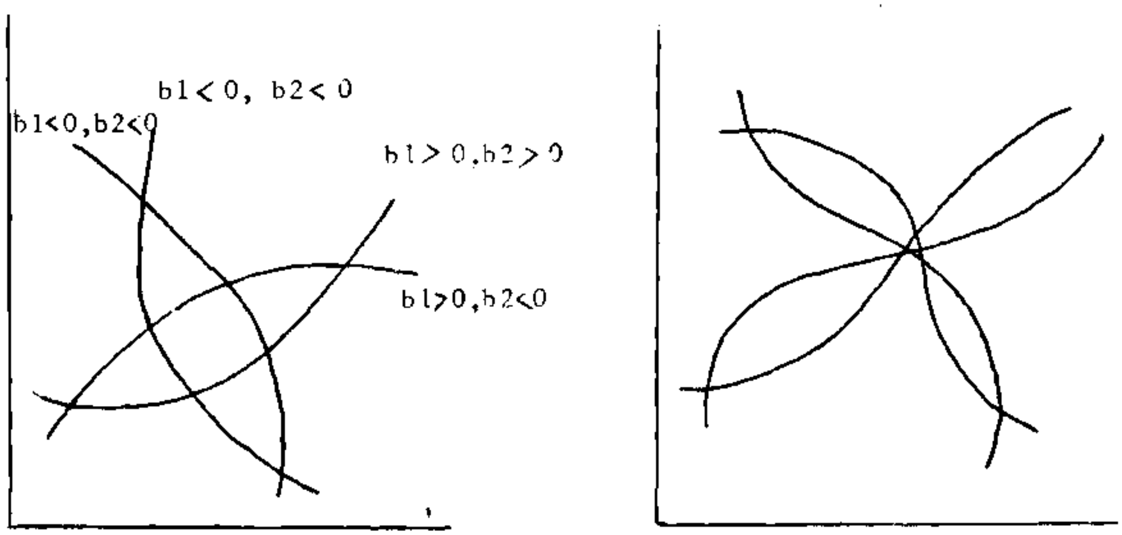
Reciprocal Function $Y = a + \frac{b}{x}$

Exponential Function $Y = a.b^{x}$

Figure 6. Curvilinear Regression Equations And Their Shape (Contd.)



Power Function $Y = a.x^b$



$Y = a + b1.X + b2.X^2$ $\qquad$ $Y = a + b1.X + b2.X^2 + b3.X^3$

Polynomial Functions

Figure 7. Curvilinear Regression Equations and Their Shapes (continued)

By making some simple transformations to each of these equations it is possible to convert them to a linear form and then all computational procedure can proceed as previously described. Let us individually describe the characterstics of each.

Reciprocal function $y = a + b / x$

Transforming the variable so that $u = 1 / x$; the equation reads

$$Y = a + b * u$$

The suitability of this function may be confirmed by plotting a scatter diagram of Y versus the trasformed variable u. If the scatter diagram takes on a linear pattern, then the linear computation may proceed by substituting $u = 1/x$

Exponential Function $Y = a * b ** X$

Transforming the variable by taking log of both sides of equation, we have

$$\log Y = \log a + (\log b)\ (X)$$

$$\text{or } Y1 = a1 + b1 * X$$

where $Y1 = \log Y$

$a1 = \log a$

$b1 = \log b$

Again the suitability of this function may be explored by plotting X versus transformed variable Y1. If the scatter takes on a linear pattern, then the computations on the transformed variable may proceed as before.

It must be remembered that the constants must be converted back to a and b by taking the anti-logs of a1 and b1. The standard error of regresssion will be in log Y units, so it is a multiplicative factor in Y units.

Power Function Y = a X ** b

Again taking the log of both sides of the equation, we have

$$\log Y = \log a + \log X$$

A double transformation is necessary here, where we use $v = \log Y$, $u = \log X$ and $a1 = \log a$

$$v = a1 + b\ u$$

When equation is obtained, a1 must be converted back to "a" by taking anti-log. Again the standard error of regression is in log Y units.

The scatter diagram would indicate which form to be used. One that gives the least standard error is the best fit.

An alternative procedure is to use the polynomial function of general form

Y = a + b1X + b2X**2 + b3X**3 + ... + bnX**n

It is possible but not practical to go beyond a cubic equation.

The general quadratic form is shown below

Y = a + b1 X + b2 X**2

The constants a, b1 and b2 may be obtained by solving simultaneously the three normal equations obtained by the same least square criterion :

Y = na + b1 X + b2 XX

XY = a X + b1 XX + b2 X**3

XXY = a XX + b1 X**3 + b2 X**4

Or the cubic form

Y = a + b1 X + b2 X**2 + b3 x**3 can be obtained by solving 4 simultaneous equations :

Y = na + b1 X + b2 XX + b3 X**3

XY = a X + b1 XX + Bb2 X**3 + b3 X**4

```
XXY = a  XX + b1  X**3 + b2  X**4 + b3  X**5

X**3Y = a  X**3 + b1  X**4 +b2  X**5 + b3  X**6
```

6.6 SPSS PROCEDURE FOR CURVILINEAR REGRESSION

As described earlier, curvilinear regression can be activated by either transformation of equation to linear model or by using polynomial function, starting with lowest order. For example,

$$Y = a + b / X$$

can be transformed as $Y = a + b\,v$ where $v = 1/X$.

The value v can be easily computed with the help of COMPUTE procedure card and then the regression coffleclents are obtained in the similar fashion as already described in linear and multilinear regression.

CURVILINEAR EQN	TRANSFORMED EQN	PROCEDURE CARD
Y = a + b / x	Y = a + b X1	COMPUTE; X1=1/X
Y =a*b**X	Y1 = a1 + b1*X	COMPUTE; Y1=lg10(Y)
Y=a*X**b	Y1=a1 + b*X1	COMPUTE;Y1=lg10(Y)
		COMPUTE; X1=lg10(X)
y=a+bX+cX**2+dX**3	Y=a+bX+cX2+dX3	COMPUTE; X2=X*X
		COMPUTE; X3=X**3

After computing these new variables the coefficient of regression equation can be computed by following procedure card

```
REGRESSION;VARIABLES=Y,X,X2,X3/
          REGRESSION=Y WITH X,X2,X3/
```

It is possible to fit a polynomial to pass thru any number of points by using higher order polynomials but then the relationship would be so complicated, and possibly unstable, that it will lose practical significance. Hence it is wiser not to go beyond cubic equation. Appendix L describes the curvilinear regression procedure cards and results.

CHAPTER VII

QAS - A SIMPLIFIED APPROACH TO REGRESSION ANALYSIS

7.1 QAS for linear regression

Many practical situations, met in industry and elsewhere can be represented by sets of equations linking a collection of relevant quantities. It is then useful to be able to manipulate data in order to bring out further quantitive information about the real situation or of the consequences of choosing particular values for the properties. The matter is complicated because of the variety of parameters which may be stated and the questions which may be asked. The very high level computer language Question Answering System on mathematical models (QAS) is a simple but powerful way of handling this type of situation.

An investigation of the model and practical exploitation of the knowledge represented by the model usually involves a resolution of the set of equations after substitution from a subset of properties with known or given values (the input subset) in order to obtain the values of some or all of the other properties (the output subset). The QAS provides convenient means for storing the algebraic models and for their resolution from any conceivable corner. The main power of QAS is in the free manipulation of input and output subsets.

In general, the QAS language as currently implemented on the Cyber 74 computer is not very convenient for statistical computation the latter usually involves fixed sets of input and output data and little use can be made of the QAS flexibility for their manipulation, QAS statistical models are rather robust because the QAS syntax does not allow subscripted variables etc. However, some QAS models for statistical calculation were found useful and convenient like, for instance, model STAT3 for generalized two-parameter linear regression with a choice of nine function types.

An example of QAS conversion is given in Appendix N. The model handles up to eight points. Note that the transformed coordinates of the points 5, 6, 7, 8 have to be assigned equal to 0, so that the whole model may be resolved by built-in standard consecutive substitution procedure.

It is not surprising that x-reciprocal relationship gives the value of correlation coefficient closest to 1 and the best fit: there is a linear relation between linear density of yarn and reciprocal value of yarn count, and also approximate linear relationships between yarn density and yarn strength as well as yarn strength and fabric strength.

7.2 QAS for curvilinear regression

The superiority of QAS for curvilinear regression over other computing techniques would be evident from the example described in Appendix O. Least Square Method has been used for regression analysis in SPSS. But the coefficient of a complicated exponential curve as shown below can not be computed with the help of SPSS because there is no way to transform the equation of a given shape to linear form like :

P = (1-a) * e**bt + a

A close look at the illustration given in Appendix P would reveal that QAS is an advanced computer language. It is extremely easy to create QAS models and to find the coefficients of more complicated forms of curvilinear relationships. The values of coefficients a, b, c, ... giving minimum sum of the squares of difference (YE-Y) have to be found, for example, by direct optimization as in Appendix P (using Nelder-Mead algorithe of Simplex optimization method available in QAS). Alternatively, we may define and put equal to zero partial derivatives of sum of squares with respect to a, b, c, ... and solve the model as a system of nonlinear equation by means of Newston-Raphson procedure, also built-in in the QAS.

CHAPTER VIII

POCKET CALCULATORS AND MICROCOMPJTERS FOR STATISTICAL ANALYSIS

We live in an era of pocket calculators, programmable pocket calculators and microcomputers. In order to assess the use of pocket calculators in statistical computation, we examined the capability of a few of them in the price range between $ 20 and several hundreds.

Texas Instruments TI-30 [69] is a cheap non-programmable calculator with one memory location. It can be used for evaluation of mean and standard deviation in rather laborious and inefficient way. Great disadvantage : no hard copy output to keep track of the user's actions and calculator's responses.

Texas Instruments SR-51-II differ from the previous one by having three memories and several special keys for initiation of the procedure of automatic evaluation of mean, standard deviation and variance as well as intercept, slope and correlation in linear regression. The disadvantages remain basically the same.

Texas Instruments TI 58/59 [68] are programmable calculators with up to 100 registers for storing the data and up to 960 registers for storing the program. There are interchangeable solid state modules available, some of them containing preprogrammed statistical procedures. The user can also create and enter his own programs, which, in case of TI-59 can be written and read from magnetic cards. The programmability or program modules reduce considerably the "key-stroking" effort needed for performing complicated routine calculations including the statistical ones. The convenience in using these calculators is further enhanced by an optional printing device providing hard copy traces of user's actions and calculator's responses (results and simple pre-programmed messages) on a narrow strip of paper. The disadvantages which still remain are the clumsiness of writing and inserting the programs using rather low-level "machine oriented" set of instructions and storage limitations.

Hewlett-Packard HP-97 programmable printing calculator [70] offers much the same facilities as the TI 58/59 including 26 data storage registers, 224 steps of program memory three levels of subroutines, four flags, 20 easily-accessed labels, standard packs with prerecorded programs and magnetic card for recording and reading user's own programs.

The pocket calculators in general offer inexpensive, portable and easy-to-use means for performing simple statistical calculations like means and variances of sample populations and two-variable regression analysis. Some more complicated statistical procedures may be programmed, but their complexity being restricted by limited storage capacity. Currently available pocket calculators should contribute significantly to developing the taste for statistical calculations among students, engineers and researchers. However, they may become a mixed blessing in serious quality control and research work in two particular respects : testing hypothesis and obscuring the results. The latter is due to limitations concernig the form and format of output. The former seems to be more serious. The capacity of pocket calculators is insufficient for providing a variety of characterstics needed for testing of hypotheses including the generating of relevant values of normal, T, F or chi-square distribution. Another practical drawback is the impossibility to store, edit and recall large collections experimental data. This all justifies usage of computers and program packages like SPSS for statistical analyses in years to come.

On the contrary, the microcomputers are personal computers with price tags not too far above pocket calculators, and information processing power and convenience not too far below the large computer installations. The microcomputers are new and their software suport is not too advanced. To the best of our knowledge there is no statistical package comparable to SPSS, available and implemented on the microcomputers yet, but there is no reason why one should or could not be. An important advantage of microcomputers is that they usually can be operated both as stand-alone computers and as terminals to large time sharing systems. It is also possible to exchange programs and data between microcomputers and large computers. In information processing networks consisting of large computers and microcomputers, the former may be used primarily for storing large quantities of data and program libraries and performing complex information processing tasks whereas the latter are more efficient in gathering data at local level and performing simpler computation (all the examples in this thesis probably all the statistical methods covered by SPSS belong to this category). It is obvious that this arrangement would open new prospects on possibilities in computer aided decision making based on statistical analysis of experimental data and previous experience. When the problem (i.e. objective function and constraints) cannot be formulated strictly in terms of linear equations and inequalities, more complicated methods of nonlinear and dynamic programming have to step in.

CHAPTER IX

CONCLUSIONS

SPSS program package was originally written for Social Sciences but it can be conveniently used in textiles and other engineering fields. Although there are a multitude of reasons for wanting to use SPSS system, an unfortunate percentage of those desiring to do so are intimidated by the size and apperent complexity of SPSS. Mathematical statistics plays an indispensible role in textile research and quality control. Use of computers in statistical analysis makes it possible to utilize the most sophisticated statistical characterstics and procedures without getting involved in laborious manual calculations, consequently to improve the design of experiment and to extract maximum information from the measurements and observations.

The "Statistical Package for Social Sciences" (SPSS) available at Georgia Tech Cyber 74 computer suits well the needs of the researcher at the School of Textile Engineering. He can use the package for measures of central values and dispersion, analysis of variance, regression analysis etc, without getting involved in writing and debugging his own programs.

This compact report will hopefully enable the user to get interested in SPSS and learn its essential steps. Only the most frequently used statistical methods and SPSS procedures are described along with several examples. There are numerous sophistications employed in SPSS to make it very efficient. But those have been avoided in this report in order to keep it compact and comprehensive. An industrious programmer may wish to obtain greater efficiency by referring to SPSS Manual. But more often than not, the casual user does not wish to spend too much time in learning a whole programming technique. This report should help this group of users.

The use of SPSS reduces required programming skills on the part of the user, but it does not substitute for the knowledge of basic principles of mathematical statistics. On the contrary, better knowledge of statistical concepts and methods, as well as, better understanding of the nature of the object of investigation will enable the user to take full advantage of the SPSS.

The thesis also indicates those areas of statistical analyses which are not covered by SPSS or for which the use of SPSS may be unnecessary and inefficient. Special computer program or QAS may help in the first case and pocket calculators in the second.

CHAPTER X

RECOMMENDATIONS

It is suggested that a user should bear the following points for running SPSS subprograms successfully.

1. The first task is to prepare a data file which is to be analysed statistically. It is assumed that the user is acquainted with the preparation of a new file on a computer. The beginner is recommended to refer to Cyber '74 Manual to learn appropriate commands to make a file.

2. The data can be arranged in Fixed or Free format.

3. The next step after making data file is to make a Data Definition File as explained in Chapter II. It consists of data definition cards like FILE NAME, RUN NAME, VARIABLE LIST, INPUT MEDIUM, INPUT FORMAT, N OF CASES, READ INPUT CARD and then a procedure name and finally EXECUTE card.

4. If Fixed format is used, always remember to use the element Fw.d irrespective of whether the numbers are integer or real.

   INPUT FORMAT;FIXED(Fw.d)

where F indicates that the variable is numerical, w indicates the column width and d is the number of digits to the right of the decimal point. SPSS does not recognize the element Iw even for integer numbers.

5. The data definition cards and procedure cards must be correctly spelled. Any change in these results in an error message.

6. The user should refer to Cyber 74 SPSS Manual for interpretation of error messages.

7. Data definition file can be edited in Auto-mode after entering SPSS, as explained in Chapter II. Any card can be corrected by using the same procedure number and rewriting the correct statement. The same number in the same file, replaces the old statement by the new one.

8. No two cards should have the same number unless replacement of the previous statement is desired.

APPENDIX A

Central Tendencies And Dispersion Values

Data file

36

37

36

32

36

37

33

34

32

32

36

38

33

39

36

39

36

37

32

31

```
AT,SPSSONL/UN=LIBRARY

/G,TEST1

/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  TEST1

AUTO-MODE

?  5.0 FILE NAME;TESTS1

?  10.0 RUN NAME;DESCRIPTIVE STATISTICS

?  15.0 VARIABLE LIST;STRENTH

?  20.0 INPUT MEDIUM;DISK

?  25.0 INPUT FORMAT;FREEFIELD

?  30.0 N OF CASES;20

?  35.0 FREQUENCIES;GENERAL=ALL

?  40.0 OPTIONS;7

?  45.0 STATISTICS;ALL

?  50.0 READ INPUT DATA

?  EXECUTE
```

MEAN	35.100	STD ERR	.561	MEDIAN	35.833
MODE	36.000	STD DEV	2.511	VARIANCE	6.305
KURTOSIS	-1.232	SKEWNESS	-.136	RANGE	8.000
MINIMUM	31.000	MAXIMUM	39.000	SUM	702.000
C.V.PCT	7.154	.95 C.I.	33.925	TO	36.275

APPENDIX B

Frequency Distribution

```
/SPSSONL

USE ANS SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  TEST1

AUTO-MODE

?  5.0 FILE NAME;TESTS1

?  10.0 RUN NAME;DESCRIPTIVE STATISTICS

?  15.0 VARIABLE LIST;STRENGTH

?  20.0 INPUT MEDIUM;DISK

?  25.0 INPUT FORMAT;FREEFIELD

?  30.0 N OF CASES;20

?  35.0 FREQUENCIES;GENERAL=ALL

?  40.0 STATISTICS;ALL

?  45.0 READ INPUT DATA

?  EXECUTE

     - - - FREQUENCIES - - -


     STRENTH
```

CODE	FREQ	REL P	ADJ P	CUM P
31	1	5.0	5.0	5.0
32	4	20.0	20.0	25.0
33	2	10.0	10.0	35.0

34	1	5.0	5.0	40.0
36	6	30.0	30.0	70.0
37	3	15.0	15.0	85.0
38	1	5.0	5.0	90.0
39	2	10.0	10.0	100.0
TOTAL	20	100.0	100.0	100.0

- - - FREQUENCIES - - -

STRENTH

MEAN	35.100	STD ERR	.561	MEDIAN	35.833
MODE	36.000	STD DEV	2.511	VARIANCE	6.305
KURTOSIS	-1.232	SKEWNESS	-.136	RANGE	8.000
MINIMUM	31.000	MAXIMUM	39.000	SUM	702.000
C V PCT	7.145	.95 C.I.	33.925	TO	36.275

- -

APPENDIX C

Histogram

```
/AT,SPSSONL/UN=LIBRARY

/SPSSONL

USE AN SPSS FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  TEST1

AUTO-MODE

?  5.0 VARIABLE LIST;STRENTH

?  10.0 INPUT MEDIUM;DISK

?  15.0 INPUT FORMAT;FREEFIELD

?  20.0 N OF CASES;20

?  30.0 FREQUENCIES;GENERAL=ALL

?  40.0 READ INPUT DATA

?  EXECUTE
```

```
ENTERING SPSS

- - - FREQUENCIES - - -

STRENTH


      I
31.   ****** ( 1)
      I
      I
32.   ********************* ( 4)
      I
      I
33.   *********** ( 2)
      I
      I
34.   ****** ( 1 )
      I
      I
36.   ******************************* ( 6 )
      I
      I
37.   **************** ( 3 )
      I
      I
38.   ****** ( 1)
      I
      I
39.   *********** ( 2 )
      I
      I
      I*********I*********I*********I*********I*********I
      0         2         4         6         8        10
```

APPENDIX D

T - Test

Data file

```
6.4,7.2
6.5,6.8
6.3,6.5
6.1,6.7
6.9,6.8
6.3,6.7
6.9,7.2
6.2,7.4
6.1,7.0
6.3,7.7
```

```
/SPSSONL
SPSSONL/ONLINE V4.0
USE AN SPSS FILE THIS RUN
?  NO
ENTER LOCAL DATA FILE NAME
?  TEST2
AUTO-MODE
?  5.0 VARIABLE LIST;ELONG1,ELONG2
?  10.0 INPUT MEDIUM;DISK
?  15.0 INPUT FORMAT;FREEFIELD
?  20.0 N OF CASES;10
?  25.0 T-TEST;PAIRS= ELONG1 WITH ELONG2
?  30.0 READ INPUT DATA
```

```
? EXECUTE

ENTERING SPSS

- - - T-TEST - - -

                ELONG1          ELONG2

NO.OF CASES     10              10

MEAN            6.4000          7.0000

STD DEV          .2906           .3712

STD ERROR        .0919           .1174

T-VALUE         -4.02

DF              18

SIG.            .001
```

The significance value of .001 means that there is 99.9 % probability of the two samples being from different populations.

APPENDIX E

Oneway Analysis Of Variance

Data file

1,56

1,55

1,62

1,59

1,60

2,64

2,61

2,50

2,55

2,56

2,45

3,46

3,45

3,39

3,43

4,42

4,39

4,45

4,43

4,41

```
/AT,SPSSONL/UN=LIBRARY

/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  TEST3 AUTO-MODE

?  5.0 FILE NAME;SIMPLE ONE WAY ANOVA

?  10.0 VARIABLE LIST;SPEED,BREAKS

?  15.0 INPUT MEDIUM;DISK

?  20.0 INPUT FORMAT;FREEFIELD

?  25.0 N OF CASES;20

?  30.0 READ INPUT DATA

?  35.0 ONEWAY;BREAKS BY SPEED(1,4)/

?  35.10 CONTRAST=1,0,0,-1

?  35.15 CONTRAST=0,1,-1,0/

?  35.20 CONTRAST=1,-1,-1,1/

?  EXECUTE
```

ENTERING SPSS

- - - ONEWAY - - -

VARIABLE BREAKS BY SPEED

ANALYSIS OF VARIANCE

SOURCE	D.F.	S.S.	M.S.	F RATIO	F PROB.
BETWEEN GROUPS	3	1135.00	378.33	29.790	.000
WITHIN GROUP	16	203.20	12.70		
TOTAL	19	1338.20			

CONTRAST COEFFICIENT MATRIX

	GRP 1	GRP 2	GRP 3	GRP 4
CONTRAST 1	1.0	0	0	-1.0
CONTRAST 2	0	1	-1	0
CONTRAST 3	1	-1	-1	1

	VALUE	S.E.	T-VALUE	D.F.	T-PROB
CONTRAST 1	16.4	2.254	7.276	16	.000
CONTRAST 2	13.6	2.254	6.034	16	.000
CONTRAST 3	-0.4	3.187	-.125	16	.902
CONTRAST 2	0	1.0	-1.0	0	
CONTRAST 3	1.0	-1.0	-1.0	1.0	

APPENDIX F

2^2 - Factorial Analysis Of Variance

Data file

```
1,1,22
1,1,16
2,1,19
2,1,23
1,2,8
1,2,13
2,2,16
2,2,15
1,3,9
1,3,10
2,3,13
2,3,16
```

```
/SPSSONL
SPSS/ONLINE V4.0
USE AN SPSS SYSTEM FILE THIS RUN
?  NO
ENTER LOCAL DATA FILE NAME
?  TEST4
AUTO-MODE
?  5.0 VARIABLE LIST;CONC,DYESTUFF,DEPTH
?  10.0 INPUT MEDIUM;DISK
?  15.0 INPUT FORMAT;FREEFIELD
?  20.0 N OF CASES;12
```

```
? 25.0 READ INPUT DATA
? 30.0 ANOVA;DEPTH BY CONC,DYESTUFF(1,3)
? EXECUTE
```

ENTERING SPSS

\- - - ANOVA - - -

DEPTH BY CONC, DYESTUFF

SOURCE	S.S.	DF	M.S.	F	F PROB.
CONC	12.500	1	12.500	0.502	.492
DYESTUFF	103.000	2	51.500	2.069	.169
CONC * DYESTUFF	70.333	2	35.167	1.493	.263
EXPLAINED	185.883	5	37.167	1.493	.263
RESIDUAL	298.667	12	24.889		
TOTAL	484.500	17	28.500		

Here the RESIDUAL is synonymous to the term ERROR used in the text. It is the difference of Total and Explained variation. Explained variation is sum of the effects of individual variables. The sign "*" between the variables as seen above, indicates interaction between them.

APPENDIX G

2^3- Factorial Analysis Of Variance

Data file

1,1,1,3.5

1,1,1,3.2

1,1,2,2.4

1,1,2,2.8

1,1,3,2.9

1,1,3,2.7

1,2,1,2.1

1,2,1,2.3

1,2,2,3.1

1,2,2,3.1

1,2,3,3.4

1,2,3,3.3

2,1,1,2.5

2,1,1,2.4

2,2,2,3.2

2,1,2,3.1

2,1,3,3.6

2,1,3,3.3

2,2,1,2.8

2,2,1,2.7

2,2,2,3.4

2,2,2,3.5

2,2,3,3.2

```
/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL FILE NAME

?  ANOVA4

AUTO-MODE

?  5.0 VARIABLE LIST;MIXING,FEED,NEEDLES,NEPS

?  10.0 INPUT MEDIUM;DISK

?  15.0 INPUT FORMAT;FREEFIELD

?  20.0 N OF CASES;24

?  25.0 READ INPUT DATA

?  30.0 ANOVA;NEPS BY MIXING,FEED,NEEDLES(1,3)

?  EXECUTE

ENTERING SPSS

* * * ANOVA * * *

NEPS BY MIXING, FEED, NEEDLES
```

SOURCE OF VARIATION	S.S.	DF	M.S.	F	F PROB.
MIXING	.224	1	.224	8.490	.013
FEED	.031	1	.031	1.173	.300
NEEDLES	1.266	2	.633	23.991	.001
MIXING * FEED	.063	1	.063	2.387	.148
MIXING * NEEDLES	.325	2	.163	6.158	.014
FEED * NEEDLES	.744	2	.372	14.097	.001
MIXING * FEED * NEEDLES	1.223	2	.611	23.169	.001
EXPLAINED	3.983	11	.362	13.722	.001
RESIDUAL	.317	12	.026		
TOTAL	4.300	23	.187		

APPENDIX H

FORTRAN PROGRAM FOR CHI-SQUARE TEST

```
      PROGRAM CHI(INPUT,OUTPUT)
      DIMENSION M(25)
101   ISUM=0
      TBREK=0
      PRINT*,"WRITE FREQUENCY VALUES, PRINT 999 AT END"
      DO 5 N=1,25
      PRINT*,N
      READ*,M(N)
      IF (M(N).EQ.999) GO TO 6
      ISUM=ISUM+M(N)
5     TBREK=TBREK+((N-1)*M(N))
6     AMEAN=TBREK/FLOAT(ISUM)
      EF1=0
      NFACT=1
      CHISQ=0
      DO 10 I=1,N
      IF(I.GT.1)NFACT=NFACT*(I-1)
      PROB=(AMEAN)**(I-1)/(FLOAT(NFACT)*EXP(AMEAN))
      EXFRQ=(FLOAT(ISUM))*PROB
      IF(EXFRQ.GE.5) GO TO 100
      EXFRQ=EXFRQ+EF1
      M(I)=M(I)+M(I-1)
      IF(I.NE.(N-1)) GO TO 102
100   CHISQ=CHISQ+(M(I)-EXFRQ)**2/EXFRQ
102   EF1=EXFRQ
```

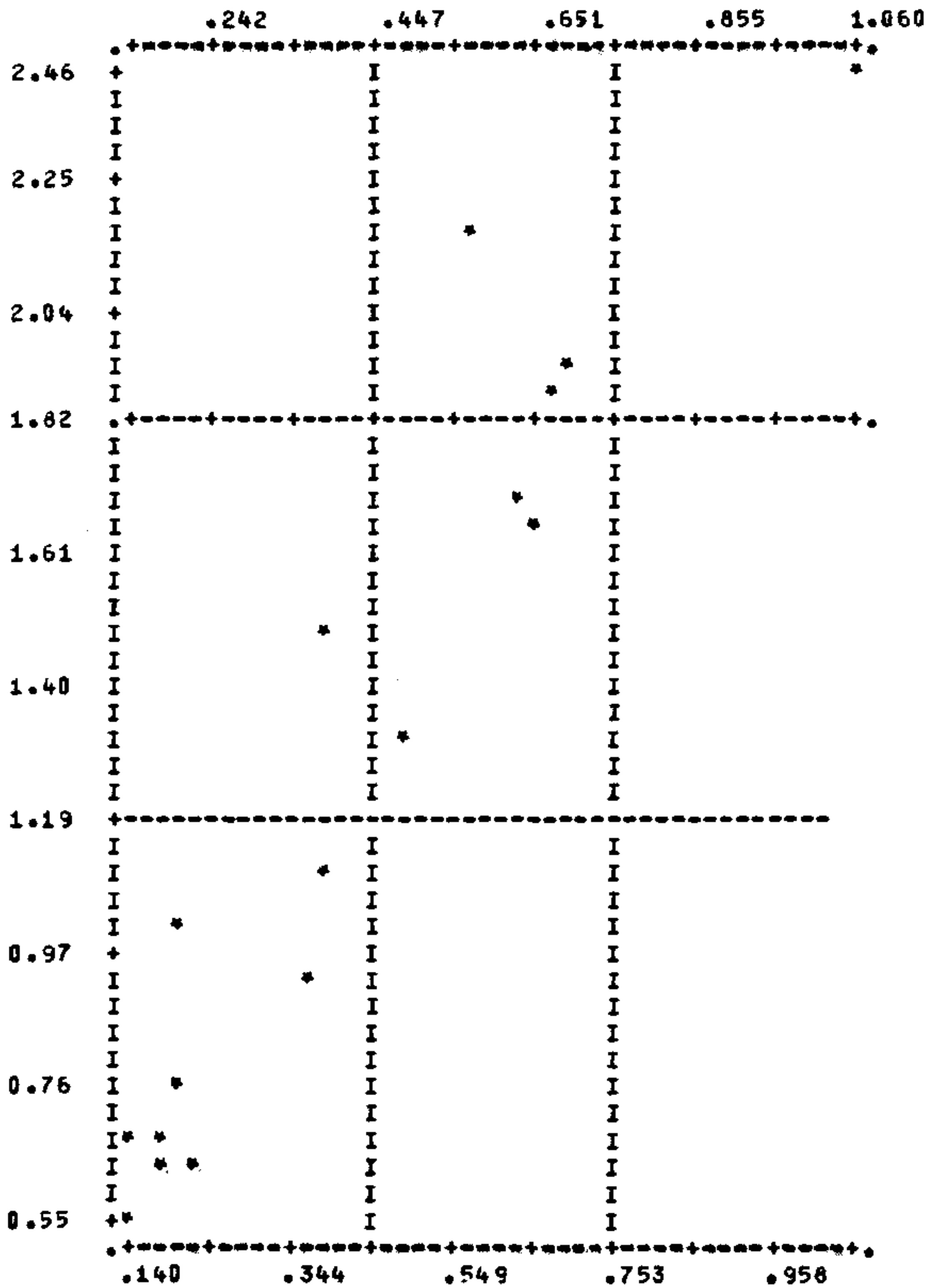
```
10    CONTINUE
      PRINT*,CHISQ
      GO TO 101
      END
```

APPENDIX I

SCATTER DIAGRAM AND PEARSON CORRELATION

```
/AT,SPSSONL/UN=LIBRARY

/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  SCATER

AUTO-MODE

?  5.0 VARIABLE LIST;RVALUE,PHEIGHT

?  10.0 INPUT MEDIUM;DISK

?  15.0 INPUT FORMAT;FREEFIELD

?  20.0 N OF CASES;18

?  25.0 READ INPUT DATA

?  30.0 SCATTERGRAM;RVALUE WITH PHEIGHT

?  35.0 PEARSON CORR;RVALUE WITH PHEIGHT

?  EXECUTE
```

- - - SCATTERGRAM - - - DOWN RVALUE, ACROSS PHEIGHT



PLOTTED VALUES - 18

EXCLUDED VALUES - 0

MISSING VALUES - 0

- - - PEARSON CORRELATION - - -

VARIABLE PAIR	COEFF.	N	SIG.
RVALUE PHEIGHT	.9458	18	.000

APPENDIX J

Linear Regression

```
/AT,SPSSONL/UN=LIBRARY

/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  SCATER

AUTO-MODE

?  5.0 VARIABLE LIST;RVALUE,PHEIGHT

?  10.0 INPUT MEDIUM;DISK

?  15.0 INPUT FORMAT;FREEFIELD

?  20.0 N OF CASES;18

?  25.0 READ INPUT DATA

?  30.0 REGRESSION;VARIABLES=RVALUE,PHEIGHT

?  30.10 REGRESSION=RVALUE WITH PHEIGHT(2)

?  EXECUTE
```

ENTERING SPSS

- - - REGRESSION - - -

DEP. VAR. --- RVALUE

MEAN RESPONSE		1.3156			
STD. DEV.		.6028			
MULTIPLE R		.9458			
R SQUARE		.8945			
STD DEV		.2018			
ADJ R SQ		.8879			
ANOVA	DF	S.S	M.S.	F	F PROB.
REGRESSION	1	5.525	5.525	135.621	.000
RESIDUAL	16	.652	.041		

VARIABLE	B	S.E. B	F	SIG.
PHEIGHT	2.159	.185	135.621	.000
CONSTANT	.355	.095	13.901	.002

APPENDIX K

FORTRAN program for creating SPSS data file from the marker bank in Table 7.

```
      PROGRAM AREAS(INPUT=65/72,OUTPUT,TAPE7,TAPE8)
C     SPSS DATA FILE GENERATOR :
C     X1=A   X2=IS   A3=JS
  1   READ (7,101) A, B
      A=1.6388888889*A*B
      IS=JS=0
  2   READ (7,*) I,K
      IF (K.LE.0) GO TO 3
      IS=IS+K*(I/100)
      JS=JS+K*MOD(I,100)
      GO TO 2
  3   WRITE (8,102) A,FLOAT(IS),FLOAT(JS)
      IF (K.EQ.0) GO TO 1
      STOP
 101  FORMAT (18X,F5.2,1X,F5.2)
 102  FORMAT (F8.2,2F6.0)
      END
```

SPSS Data File Representing Area, Length and Waist

365.27	70.	64.	342.51	63.	64.	1436.88	306.	209.	1464.24	302.	220.
337.68	69.	56.	327.60	67.	55.	1473.05	306.	219.	1356.68	268.	234.
369.82	74.	61.	373.10	69.	67.	1523.89	304.	238.	1357.59	269.	232.
325.66	63.	59.	349.76	64.	65.	1465.28	301.	223.	1850.87	373.	299.
361.54	68.	65.	363.63	72.	61.	1513.77	311.	227.	1877.84	384.	297.
353.57	68.	62.	367.99	73.	61.	1442.51	297.	223.	1848.85	363.	311.
372.12	68.	68.	356.56	68.	63.	1452.31	301.	218.	1895.67	368.	318.
358.90	64.	68.	353.50	70.	60.	1486.45	309.	223.	1867.29	364.	313.
350.66	67.	62.	349.04	69.	60.	1437.13	305.	212.	2079.43	435.	324.
341.94	67.	60.	347.95	70.	58.	1477.17	306.	223.	2129.71	432.	337.
370.80	73.	62.	332.88	67.	57.	1503.69	309.	228.	2129.56	428.	343.
348.84	69.	60.	353.36	68.	62.	1476.69	311.	218.	2105.39	428.	334.
601.09	66.	68.	348.85	70.	59.	1422.49	307.	206.	2125.67	420.	349.
342.12	72.	54.	354.99	70.	60.	1425.42	307.	206.	2132.06	428.	341.
353.69	68.	62.	338.95	68.	59.	1454.	309.	219.	2170.92	436.	344.
370.61	67.	68.	354.67	71.	59.	1420.61	305.	208.	2134.27	420.	348.
342.39	65.	62.	374.60	70.	66.	1437.55	304.	211.	2508.93	508.	397.
316.94	71.	61.	380.74	71.	67.	1436.61	319.	216.	2450.08	507.	385.
354.66	67.	64.	384.82	71.	68.	1514.22	317.	225.	2504.62	505.	401.
360.56	64.	68.	375.62	71.	66.	1454.30	308.	212.	2542.93	506.	410.
341.38	71.	55.	698.76	129.	135.	1466.73	305.	218.	2531.24	504.	406.
326.59	62.	60.	744.79	138.	133.	1445.11	309.	210.	2476.30	507.	391.
336.92	73.	52.	1084.74	210.	178.	1448.40	308.	211.	2492.32	508.	392.
352.34	72.	58.	1096.29	214.	187.	1459.96	309.	214.	2574.30	520.	405.
362.66	71.	62.	1475.32	306.	220.	1461.69	306.	215.	2512.28	507.	399.
328.04	62.	61.	1374.27	287.	211.	1495.16	310.	223.	2455.31	509.	382.
326.79	65.	57.	1377.36	286.	212.	1446.26	302.	216.	2455.50	508.	382.
344.55	64.	64.	1379.09	285.	212.	1486.67	307.	223.	2511.84	506.	401.

Multiple Regression

```
AT,SPSSONL/UN=LIBRARY

/SPSSONL

SPSS/ONLINE V4.0

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

?  MARKER

AUTO-MODE

?  5.0 VARIABLE LIST;AREA,LENGTH,WAIST

?  10.0 INPUT MEDIUM;DISK

?  15.0 INPUT FORMAT;FREEFIELD

?  20.0 N OF CASES;113

?  25.0 READ INPUT DATA

?  30.0 REGRESSION;VARIABLES=AREA,LENGTH,WAIST

?  30.10 REGRESSION=AREA WITH LENGTH,WAIST

?  EXECUTE
```

ENTERING SPSS.

- - - REGRESSION - - -

DEP. VAR. --- AREA

MEAN RESPONSE	1154.0485
STD DEV	777.6345
FINAL STEP.	
MULTIPLE R	.9995
R SQUARE	.9990
STD DEV	24.6271
ADJ R SQR	.9990

ANOVA	DF	S.S.	M.S.	F	SIG.
REGRESSION	2	.68E+08	.33E+08	.56E+05	.612
RESIDUAL	110	.68E+05	606.50	.000	

VARIABLE	B	S.E.	F	SIG.
LENGTH	2.979	0.114	679.063	.000
WAIST	4.308	0.152	269.342	.000
CONSTANT	-.008	4.308	.000	.999

APPENDIX L

Multiple Regression And Fabric Geometry

KC	KW	SC	SW	K
4.036	4.557	4	5	13.492
4.448	4.033	3	7	15.063
5.884	3.355	1	7	18.065
4.171	4.459	3	5	13.207
4.837	3.721	1	8	16.126
5.894	3.180	0	9	18.117
3.778	4.644	2	9	11.828
4.497	3.763	1	8	14.491
5.423	3.482	1	11	16.298
4.081	4.883	4	2	13.550
4.291	4.617	4	3	14.362
4.870	4.117	5	5	15.950
4.273	4.550	4	3	14.011
4.554	4.250	4	2	15.029
4.791	4.050	3	4	15.613
3.669	4.531	2	5	11.982
4.032	3.984	1	5	14.245
4.641	3.452	1	5	16.471
4.255	4.281	5	2	13.498
4.371	4.064	5	2	14.904

4.871	3.746	3	3	15.672
4.054	4.381	5	2	13.017
4.498	3.984	4	43	14.200
5.118	3.587	4	4	15.753
3.805	4.361	2	4	12.062
4.258	3.852	1	5	13.610

```
/AT,SPSSONL/UN=LIBRARY /SPSSONL

SPSSONLINE

USE AN SPSS SYSTEM FILE THIS RUN

?  NO

ENTER LOCAL DATA FILE NAME

DREG

AUTO-MODE

?  5.0 VARIABLE LIST KC,KW,SC,SW,K

?  10.0 INPUT MEDIUM DISK

?  15.0 INPUT FORMAT FREEFIELD

?  20.0 N OF CASES 27

?  25.0 REGRESSION VARIABLES=KC,KW,SC,SW,K

?  25.10 REGRESSION=KC WITH SC,SW,K

?  25.15 REGRESSION=KW WITH SC,SW,K

?  READ INPUT DATA

?  EXECUTE
```

ENTERING SPSS

- - - REGRESSION - - -

DEP. VAR. ---- KC

MEAN RESPONSE	4.53211
STD DEV	.57591
MULTIPLE R	.9448
R SQUARE	.8936
STD DEV	.2007
ADJ. R SQ.	.8786

ANOVA	DF	S.S.	M.S.	F-VAL	F-PROB.
REGRESSION	3	7.697	2.566	63.707	0
RESIDUAL	23	.926	.040		

VARIABLE	B	S.E.	F-VAL	F-PROB.
K	0.312	.025	154.270	0
SW	0.040	.023	3.060	.094
SC	0.029	.036	.667	.423
CONSTANT	-.333	.422	.624	.437

DEP. VAR - - - KW

MEAN RESPONSE	4.04185
STD DEV	.46775
MULTIPLE R	.8963
R SQUARE	.8034
STD DEV	.2205
ADJ. R SQ.	.7777

ANOVA	DF	S.S.	M.S.	F-VAL	F-PROB
REGRESSION	3	4.570	1.523	3.323	.000
RESIDUAL	23	1.119	0.049		

VARIABLE	B	S.E.		F-
K	-.198	.028	51.367	0
SC	0.129	.040	10.590	.003
SW	0.020	.025	0.619	.440
CONSTANT	6.500	.464	196.504	0

APPENDIX M

Curvilinear Regression Analysis

Data file

```
82,10
51,8
70,13
59,10
57,12
52,6
93,12
99,12
90,10
43,8
68,10
71,13
82,14
98,14
95,13
```

```
AT,SPSSONL/UN=LIBRARY
/SPSSONL
SPSS/ONLINE V4.0
USE AN SPSS SYSTEM FILE THIS RUN
? NO
ENTER LOCAL DATA FILE NAME
```

```
?  CURVE
AUTO-MODE
?  5.0 VARIABLE LIST SCORE,SALES
?  10.0 INPUT MEDIUM;DISK
?  15.0 INPUT FORMAT;FREEFIELD
?  20.0 N OF CASES;10
?  25.0 READ INPUT DATA
?  30.0 COMPUTE SCORE2=SCORE*SCORE
?  35.0 REGRESSION VARIABLES=SALES,SCORE,SCORE2/
?  35.10 REGRESSION=SALES WITH SCORE,SCORE2/
?  EXECUTE


ENTERING SPSS.
- - - REGRESSION - - -
DEP.  VAR.  --- SALES
MEAN RESPONSE              11.0000
STD DEV                    2.3905
MULTIPLE R                 .7199
R SQUARE                   .5183
STD DEV                    1.7920
         ADJ R SQ          .4380
ANOVA          DF   S.S.      M.S.      F-VAL     F-PROB
REGRESSION     2    41.465    20.732    6.456     .012
RESIDUAL       12   38.535    3.211
```

VARIABLE	B	S.E.	F-VAL	F-PROB
SCORE	.426	.257	2.757	.123
SCORE2	-.002	.002	1.783	.207
CONSTANT	-6.962	8.949	.605	.452

APPENDIX N

QAS FOR MULTIPLE REGRESSION

Yarn count	Strength of fabric
8	1.5
16	0.71
32	0.32
36	0.30

Assuming a linear relationship first:

```
AT,QAS/UN=LIBRARY
/QAS
?  AM STAT3
?  AM REG4
?  AI X1 8,X2 16,X3 32,X4 36,XT5,XT6,XT7,XT8
?  AI Y1 1.52,Y2 .71,Y3 .3,YT5,YT6,YT7,YT82,Y4 .3
?  NP 4
FT ?  LINEAR
RESULTS:
       A1  1.6338E+00
       B1 -4.0057E-02
       CC  9.2803-01
```

The correlation coefficient is not very good, so, let us try other function types using so called Scalar model.

```
?  MS
```

```
MULTIPLE SCALAR MODE, NVP?  1

NAME?  FT

NUMBER OF CASES ?  5

1 CASE

FT ?  EXPONENT

2 CASE

FT ?  LOGARITHM

3 CASE

FT ?  POWER

4 CASE

X-RECIPROC

5 CASE

Y-RECIPROC

RESULTS:
```

CASE	A1	B1	CC
1	7.4077E-01	-5.6851E-02	9.8222E-01
2	3.0989E+00	-8.0206E-01	9.7886E-01
3	2.7003E+00	-1.0981E+00	9.9940E-01
4	-6.7920E-02	1.2663E+01	9.9973E-01
5	-1.3946E-01	9.8723E-02	9.9817E-01

APPENDIX P

QAS For Curvilinear Regression

```
/AT,QAS,CQAS/UN=LIBRARY

/QAS

MASIA - LEAST SQ METHOD FOR Y=(1-A)*EXP(B-X)+A

  15    0    5    0

 1    XE1 EXPERIMENTAL COORDINATES
 2    XE2 -------- " -------
 3    XE3 -------- " -------
 4    XE4 -------- " -------
 5    YE1 -------- " -------
 6    YE2 -------- " -------
 7    YE3 -------- " -------
 8    YE4 -------- " -------
 9    Y1 CALCULATED VALUES
10    Y2 ------- " -------
11    Y3 ------- " -------
12    Y4 ------- " -------
13    A EQUATION COEFFICIENTS
14    B EQUATION COEFFICIENTS
15    SS SUM OF SQUARES

 1    Y1=(1-A)*EXP(B*XE1)+A
 2    Y2=(1-A)*EXP(B*XE2)+A
 3    Y3=(1-A)*EXP(B*XE3)+A
 4    Y4=(1-A)*EXP(B*XE4)+A
```

```
 5     SS=(Y1-YE1)**2+(Y2-YE2)**2+(Y3-YE3)**2+(Y4-YE4)**2

MASIB - LEAST SQ METHOD FOR Y=(1-A-C*XE)*EXP(B*X)+A+C*XE
  16     0     5     0

 1     XE1
 2     XE2
 3     XE3
 4     XE4
 5     YE1
 6     YE2
 7     YE3
 8     YE4
 9     Y1
10     Y2
11     Y3
12     Y4
13     A
14     B
15     C
16     SS

 1     Y1=(1-A-C*XE1)*EXP(B*XE1)+A+C*XE1
 2     Y2=(1-A-C*XE2)*EXP(B*XE2)+A+C*XE2
 3     Y3=(1-A-C*XE3)*EXP(B*EX3)+A+C*XE3
 4     Y4=(1-A-C*XE4)*EXP(B*XE4)+A+C*XE4
 5     SS=(Y1-YE1)**2+(Y2-YE2)**2+(Y3-YE3)**2+(Y4-YE4)**2
```

The above two QAS models are very similar except that the regression equation is slightly different. The purpose is to explore which equation fits best to the example described below.

The values of strength of a NOMEX fabric, when exposed to sunshine for 50, 100, 150 and 200 hours exhibited a drop to 80.3, 69.9, 64.4 and 61.5%. The degration process generally follows the relationship

P=(1-a).e**(b.x)+a

or more generally

P = 1 - (a+ct)(e**bt=1)

What are the constants a and b which fit best the experimental values using LSM ?

```
?  DF MASIA AO A,B,SS AI XE1 50, XE2 100,XE3 150
?  AI XE4 200,YE1 .803,YE2 .699,YE3 .644,YE4 .615
AP A .5,B -1, MN SS
```

RESULTS:

```
        A     8.5080E-01
        B    -8.7992E-01
        SS    1.4725E-03
```

Now let us try the regression equation used in MASIB and see if that fits better.

```
?  DF MASIB;AO A,B,C,SS;AI XE1 50, XE2 100,XE3 150
?  AI XE4 200,YE1 .803,YE2 .699,YE3 .644,YE4 .615
AP A .5,B -1,C;MN SS
```

RESULTS:

```
        A     6.5080E-01
        B    -8.7992E-01
        C    -1.2919E-03
        SS    1.4725E-03
```

The value of SS in this case is lower than the corresponding value in MASIA, which means that the equation

P=(1-a-C.X).e**(b.X)+a+C.X

is a closer approximation to the best fit.

BIBLIOGRAPHY

1. Baggett R.B., Computer Use in Textile Mills, Textile Mercury, 140, 716-719, May 1959.

2. Borwankar D.S., The Design of Experiments for Mill Trials, Textile Institute and Industry, May 1964.

3. Brearley A., An outline of Statistical Methods used in the Textile Industry, Wool Industries Research Association, Torridon, Leeds.

4. Cooper W.D., The Co-ordination of Multiplant Textile Firms thru Mathematical Programming, Textile Institute and Industry, 1970 , 8, No 5, 132-133.

5. Enrick N.L.and Lawrence S.C., Significance Testing in Mill Experimentation, Modern Textile Magazine, 40,48-49, January 1959.

6. Enrick N.L., How to Use Variation Flow Analysis for Quality Control, Textile World, February 1962.

7. Ferrel E.B., A Median, Using Midrange Chart Run-size Subprograms, Industrial Quality Control, April 1964.

8. Hader R.C., Sampling Textile Materials, Textile Bulletin, April 1961.

9. Hicks C.R., The Analysis of Covariance, Industrial Quality Control, December 1965.

10. Harrison H., Statistical and Operational Research Techniques in the Textile Industry, Textiles of Ireland. No 7,13-15, July 1968.

11. Iscol L., How to Solve Optimization Problems, Chemical Engineering, February 1962.

12. Konopasek M and Wustrow K., Computer Assisted Engineering Design and control in Knit Fabric Production, XXII Congress of International Federation of Knitting Technologists, Oct 77, Boras, Sweden.

13. Konopasek M., An Advanced Question Answering System on sets of Algebraic Equations in Interactive Systems, ed. D. Lewin, Online Publications, Uxbridge, England, 1975.

14. Louis G.L. and Louis A.F., Practical Application of the Multiple Regression Equation, Textile Bulletin, December 1967.

15. Little H.W., Spinning How to Control out-of-control End Breakage. Textile World,98-99, September 1966.

16. Nie H. N. et al, SPSS - Statistical Package for Social Sciences, McGraw Hill Book Company, New York.

17. Newell G. and Whiteley K.J., Some Observations on the Detection of Divergent Laboratories in Round Tests, Journal of Textile Institute, No.11, 1975.

18. Verbeck A.R., A Statistical Method of Characterizing

Spinning Quality, Textile Research Journal, Vol.35, Jan. 1965.

19. Weiss H.K. , Why Use Statistical Methods, Dyestuffs, March 1959.

20. Duckworth W. E. & Wyatt J.K., Rapid Statistical Techniques for operational research workers, Operation Research Quart. 9, No. 3, 218-233 (1958).

21. Mandel J. & Lashof T. W., Interlaboratory Evaluation of Testing Methods, ASTM Bull. No. 239, 53-61, July 1959.

22. Waters W. T., Introduction to Physical Textile Testing, Alabama Polytechnic Inst., School of Textile Technology, Auburn, Alabama, 1959.

23. Murphy T. et al, Statistical Methods for Textile Technologists, (Brit. Cotton Ind. Research Assoc.). The Textile Institute, Manchester, 1960.

24. Enrick N. L., Starting a Quality Control Program, Textile Indus. 123, 70-72, January 1959.

25. Duncan A. J., Quality Control and Industrial Statistics, Richard D. Irwin Inc., Homewood, Ill., 1959.

26. Lighthill M. J., Mathematics and Technology, J. Textile Inst. 51, 408-412, August 1960.

27. Peterson E. L. and Willy N. Y., Statistical Analysis and Optimization of Systems, 1961.

28. Box G. P. E. and Jenkins G. M., Some Statistical Aspects of Adaptive Optimization and Control, J. of Royal Soc., Ser. B, 24, No. 2, 297-343, 1962.

29. Hafley W. L. and Lewis J. S., Analyzing Messy Data, Industrial Engineering Chem. 55, 37-39, April 1963.

30. Vaidya S. G., Statistical Quality Control as a Management Technique, Textile Digest 24, 43-47, April-June 1963.

31. Richter W. and Fischer K., Different Methods of Evaluating the Parameters of a Poisson Distribution, Deut. Textiltech 14, No. 1, 39-44, 1964.

32. Bicking C. A., Sequential Experiment Design for Small Samples from Successive Production, Bulletin of Intrn. Statistical Inst. 39, No. 3, 393-397, 1962.

33. Burrows G. L., Statistical Tolerance Limits - What are they ?, Appl. Statistics 12, 133-144, June 1963.

34. Savitzky A. and Golay M. J. E., Smoothing and Differentiation of Data by Simplified Least Square Procedure, Anal. Chem. 36, 1627-1639, Jul y 1964.

35. Hunt S. B., Development of Statistical Knowledge on End Uses of Textiles in the United States, Rayonne et Fibres Synthet. 20, No. 12, 1063-1074, 1964. Thru Shirley Inst. 45, 724, 1965.

36. Goodman L. A., Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables, J. Am. Statistical Assoc. 59, 319-352, June 1964.

37. Canty E. T., A Unified View of ANOVA, ASQC Tech, Conf. Trans. 19, 378-386, 1965.

38. Atkinson A. C., Statistical Designs for Pilot-Plant and Laboratory Experiments, Chem. Eng. 73, 149-154, May 1966.

39. Cochran W. G., Testing Two Correlated Variances, Technometrics 7, 447-449, August 1965.

40. Moskowitz A. D. and Cranney J. L., Multiple Linear Regression Analysis as an Aid to Work Study, (U.S. Defence Supply Center), Work Study 15, 7-11, August 1966.

41. Bartholomew D. J., Hypothesis Testing When the Sample Size is Treated as a Random Variable, J. Roy. Statistical Soc. Ser. B, 29, No. 1, 53-82, 1967.

42. Hahn G. J. and Shapiro S. S., The Use and Misuse of Multiple Regression, Ind. Quality Control 23, 181-189, Oct. 1966.

43. Booth J. E., Principles of Textile Testing: An Introduction to Physical Methods of Testing Textile Fibers, Yarns and Fabrics, London, Heywood Book, 1968.

44. Chou Yu-lun, Statistical Analysis with Business and Economic Applications, Holt, Rinehart and Winston Inc., N.Y., 1969.

45. Braverman J.D. and Stewart W.C., Statistics for Business and Economics, The Renald Press Company, N.Y., 1973.

46. Schaefer G.D., Automatic Pattern control in carpet tufting, M.S. Thesis, Georgia Institute of Technology, 1972.

47. Whitworth L.B., Alterations of Physical Properties of Long Staple Cotton by Combing, Drawing and Roving, M.S. Thesis, Georgia Institute of Technology, 1968.

48. Tehrani H.B., A Microwave Unit for the Continuous Drawing Process of Polyester-Cotton Blend Fabrics. M.S. Thesis, Georgia Institute of Technology, 1974.

49. Loti R.D., A Study of the Effect of Surfactants on the wash Fabrics of Dispersed-Dyed Fabrics, M.S. Thesis, Georgia Institute of Technology, 1972.

50. Osman L.B., Effects of Intermolecular Bonding on the Drawing Behavior of Nylon 66, M.S. Thesis, Georgia Institute of Technology, 1971.

51. Cromer E.D., The Effects of Drawing, Roving and Spinning on the Physical Properties of Empire Wr Cotton Fibers, M.S. Thesis, Georgia Institute of Technology, 19 58.

52. Chao T.K.T., A Study of Mechanics of Triaxial Fabrics, M.S. Thesis, Georgia Institute of Technology, 1976.

53. Ogletree R.K., An Investigation of the Mixing of Cotton Fibers by a Pnumatic Delay Device, M.S. Thesis, Georgia Institute of Technology, 1971.

54. Allen R.C., An Organizational and Educational Effectiveness Study of Tex-Tech Enterprise. M.S. Thesis, Georgia Institute of Tecnology, 1976.

55. Mauldin G.E., Reduced Pollution Loading from Open-Beck Dyeing of Nylon Carpets, M.S. Thesis, Georgia Institute of Technology, 1974.

56. Tata A.S., Technical Feasibility of Dyeing Nylon Carpets in a Reconstituted Dye-Bath, M.S. Thesis, Georgia Institute of Technology, 1976.

57. Knight L.L., The Effect of Particle Size and Distribution on the Optical Assessment of Soiling, M.S. Thesis, Georgia Institute of Technology, 1976.

58. Gunther D.H., An Evaluation of Fiber Friction at Low Normal Forces, M.S. Thesis, Georgia Institute of Technology, 1968.

59. Cuthburtson C.R., Interfiber Friction as Related to Fiber Diameter, M.S. Thesis, Georgia Institute of Technology, 1970.

60. Padgett H.S., Properties of Synthetic Fibers at High Strain Rates, M.S. Thesis, Georgia Institute of Technology, 1973.

61. Ko F.K.F., The Relationship Between the Fabric Construction and Tuftability of Jute Carpet Backing Fabrics, M.S. Thesis, Georgia Institute of Technology, 1971.

62. Pulls R.R., A Study of Ammendments to the Kubelka-Munk Equation, M.S. Thesis, Georgia Institute of Technology, 1972.

63. Stuart Syen, The Optimization of Open End Spinning with respect to Energy Consumption, M.S. Thesis, Georgia Institute of Technology, 1976.

64. Sakai T., Effect of Yarn Length on Tensile Strength and its Distribution, M.S. Thesis, Georgia Institute of Technology, 1972.

65. Simmons J.F., Investigation of a Minuature Spinning System as a Screening and for Spin Finish Components of Polyacrylonitrile Fibers, M.S. Thesis, Georgia Institute of Technology, 1967.

66. Bates M.R., An Investigation of the Effect of Machine and Yarn Parameters on the Strength of Tufted Jute Fabrics, M.S. Thesis, Georgia Institute of Technology, 1970.

67. Schilling E.P., An Investigation of Static Prevention in Woolen Carpets using Conductive Materials, M.S. Thesis, Georgia Institute of Technology, 1973.

68. Personal Programming, Complete Owner's Manual for Progammable 58/59, Texas Instruments Inc., 1977.

69. Calculating Better Decisions, Texas Instruments Inc., 1977.

70. HP-97, Owner's handbook in programming guide, Hewlett-Packard Company, 1975.

71. Goff H., Cut Scheduling for Optimum Fabric Utilization in Apparel Production, M.S. Thesis, M.S. Thesis, Georgia Institute of Technology, 1976.

72. Ko F.K.F., Nonlinear Viscoelasticity of Polyamide Fibers, Phd. Thesis, Georgia Institute of Technology, 1976.